# Investigation of differential gene expression in Parkinson's disease patients: A whole transcriptome approach

Genevie Borrageiro

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (Human Genetics) in the Faculty of Medicine and Health Sciences at Stellenbosch University




Supervisor: Professor Soraya Bardien
Faculty of Medicine and Health Sciences
Department of Biomedical Sciences

Co-Supervisor: Professor Sian Hemmings
Faculty of Medicine and Health Sciences
Department of Biomedical Sciences

December 2016

# DECLARATION

By submitting this thesis/dissertation, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: ……

Date: .....

# Abstract

Parkinson's disease (PD) is the most common neurodegenerative movement disorder and is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta. Dopaminergic neuronal loss results in motor symptoms such as resting tremor, bradykinesia, rigidity and postural instability. Although several PD-causing genes have been identified, the process(es) that lead to progressive neuronal loss is poorly understood. Therefore, PD research efforts have turned to genome-wide next generation sequencing technologies to provide clues to the etiology. The focus of this study was to investigate the entire transcriptome in South African patients with PD (particularly in the under-studied mixed ancestry population) to identify biological pathways that may shed light on the pathobiology of PD.

A total of 40 study participants (20 patients and 20 controls) were recruited for the study. The PD patients were recruited from the Movement Disorders clinic at Tygerberg Hospital in Cape Town, South Africa and had to meet the UK Parkinson's disease Society Brain Bank Diagnostic Criteria for PD. All individuals were from the South African mixed ancestry ethnic group which is a unique admixture of Khoisan, Black, Caucasian, and Asian populations. Copy number variation (CNV) detection in known PD genes (*SNCA*, *PARK2* (*Parkin*), *UCHL1, PINK1, PARK7, LRRK2, GCH1* and *ATP13A2* and two point mutations (*SNCA* A30P and *LRRK2* G2019S) was determined using the Multiplex Ligation-dependent Probe Amplification (MLPA) technique. Total RNA was extracted from whole blood samples and RNA-Sequencing was performed at NXT-Dx (Gent, Belgium) on the Illumina HiSeq® 4000. Bioinformatic analysis was performed using Partek® Flow® software. DAVID and Ingenuity Pathway Analysis (IPA) were used for enrichment analysis and to identify possible biological pathways involved. One candidate gene of interest was selected for further verification by quantitative real-time PCR (qPCR).

MLPA analysis revealed a heterozygous exon 2 deletion in *PARK2* in one patient however a second mutation in this gene was not identified. Therefore, all patients potentially have idiopathic PD and were analysed as one group for RNA-Seq. All samples produced good quality reads, as determined using FastQC (scores > 39) and on average 95.3% of the transcripts generated for each sample could be aligned to the reference genome (hg19). Bioinformatics analysis resulted in a candidate gene list of 132 differentially expressed genes. Analysis of these genes using IPA identified five significant dysregulated canonical pathways which included regulation of eIF4 and p70S6K signaling, EIF2 signaling, LPS/IL-1 mediated inhibition of RXR function, xenobiotic metabolism signaling and maturity onset diabetes of

young (MODY) signaling which contribute to PD pathogenesis. A possible link between *CEBPA*, *PGC-1α* and PD was also highlighted as both genes were in the prioritized candidate list and IPA gene network. *CEBPA* has been shown to interact with known PD genes and *PGC-1α* is a transcriptional regulator of mitochondrial biogenesis a key pathway linked to PD. *CEBPA* was prioritized for replication studies using qPCR and was found to be significantly down-regulated in patients which contrasted with the RNA-Seq results.

The present study revealed candidate genes, *CEBPA* and *PGC-1α* which could potentially be involved in the development of PD and a possible link to diabetes through mitochondrial mechanisms. We speculate that the increased *PGC-1α* levels observed is in response to loss of mitochondria, which leads to increased levels of reactive oxygen species that ultimately result in death of dopaminergic neurons. Our study illustrates that the use of RNA-Seq in combination with IPA is a powerful approach that may reveal candidate genes and biological pathways involved in PD. A better understanding of the molecular mechanisms underlying PD is critical for development of therapeutic modalities in order to prevent, stop or reverse dopaminergic neuronal loss in PD patients.

# Opsomming

Parkinson siekte (PD) is die mees algemene neurodegeneratiewe beweging afwyking en word gekenmerk deur die verlies van dopaminergiese neurone in die substantia nigra pars compacta. Die verlies van hierdie neurone veroorsaak motoriese simptome soos rustende rillings, bradykinesia, rigiditeit en posturale onstabiliteit. Alhoewel 'n klomp gene al geïdentifiseer was vir die oorsaak van PD is die proses(se) vat lei tot die verlies van neurone nog onduidelik. Vir die rede het navorsing studies begin fokus op genoom-wye 'next generation sequencing' tegnologie om die eitiologie van PD te verstaan. Die hoof doel van die navorsing was om die hele transkriptoom van Suid Afrikaanse patiente met PD te ondersoek om die biologiese paaie te identifiseer, met die gevolg dat meer inligting rakende die pathobiologie van PD ontbloot sal word.

'n Totaal van 40 deelnemers (20 pasiënte en 20 kontroles) is gebruik vir die studie. Die PD patiënte was gewerf van Tygerberg Hospitaal se beweging wanorde kliniek in Kaapstad, Suid-Afrika. Die patiënte moes voldoen aan die UK Parkinson's Disease Society Brain Bank Diagnostic kriteria. Alle individue was van Suid-Afrika en het bestaan uit verskeie etniese groepe, onder andere Khoisan, wit, swart en Asiatiese populasies. Deur gebruik te maak van die Multiplex Afbinding-afhanklike Probe versterking (MLPA) tegniek was aantal kopie variasie (CVN) ondek in bekende PD gene (*SNCA, PARK2 (Parkin), UCHL1, PINK1, PARK7, LRRK2, GCH1* en *ATP13A2* en in twee puntmutasies (*SNCA A30P* en *LRRK2 G2019S*). Totale RNA was onttrek van heel bloedmonsters en RNS-volgordebepaling (RNS-Seq) was gedoen deur NXT-Dx (Gent, België) op die Illumina HiSeq® 4000. Bioinformatiese analise is uitgevoer met behulp van Partek® Flow sagteware. DAVID en Ingenuity pathway analysis (IPA®) is gebruik om die analise te verryk en om moontlike biologiese paaie betrokke te identifiseer. Een kandidaat geen was gekies vir verdere analise en verifikasie deur kwantitatiewe real-time PCR (qPCR) was voltooi.

MLPA analise het in een van die patiënte 'n heterosigotiese exon 2 verlies in *PARK2* onthul, maar 'n tweede mutasie was nie in die geen geïdentifiseer nie. Dus het alle pasiënte potensieel idiopatiese PD en is ontleed as een groep vir RNS-Seq. Alle monsters het hoë kwaliteit 'reads' geproduseer en dit was bepaal deur FastQC (tellings > 39) en 'n gemiddeld van 95.3% van die transkripsies wat gegenereer was vir elke monster was in lyn met die verwysing genoom (hg19). Bioinformatika analise het gelei tot 'n kandidaat geen lys van 132 differensieel uitgedrukte gene. Ontleding van hierdie gene met behulp van IPA® het vyf belangrike dysreguleërde kanonieke paaie geïdentifiseer, wat regulering van eIF4 en

p70S6K sein, EIF2 sein, LPS / IL-1 bemiddel inhibisie van RXR funksie, xenobiotiese metabolisme sein en maturity onset diabetes of young (MODY) sein insluit wat bydrae tot PD. A moontlike skakel tussen *CEBPA*, *PCG-1*$\alpha$ en PD was tot lig gebring siende dat beide gene in die geprioritiseerde kandidaat lys en in die “IPA” netwerk voorgekom het. Vorige studies het getoon dat *CEBPA* interaksie het met bekende PD gene, terwyl *PCG-1*$\alpha$ ‘n transkripsionele reguleerder van die mitochondriale biogenese padweë is, ‘n belangrike padweg in PD. 'n Moontlike verband tussen *CEBPA, PGC-1α* en *PD* is ook uitgelig. *CEBPA* was geprioritiseer vir verdere analise met die qPCR en resultate het gewys dat die geen aansienlik minder gereguleer is in pasiënte, wat in teenstelling is met die RNA-Seq resultate.

Die huidige studie het die volgende kanidaat gene aan die lig gebring: *CEBPA* en *PGC-1α* wat albei potensieël betrokke kan wees in die ontwikkeling van PD met ‘n moontlike skakel tot diabetes deur mitokondriale meganismes. Ons spekuleer dat mitokondriale verlies die waarneemende verhoogte vlakke van *PGC-1α* veroorsaak, en hierdie kan lei tot verhoogde vlakke van reaktiewe suurstof spesies wat uiteindelik die dood van dopaminergiese neurone veroorsaak. Ons studie dui aan dat die gebruik van RNA-Seq in kombinasie met IPA® 'n kragtige benadering is, wat kandidaat gene en biologiese padweë betrokke by PD kan onthul. 'n Beter begrip van die molekulêre meganismes onderliggend aan PD is van kritieke belang vir die ontwikkeling van nuwe therapie om die verlies van dopaminergiese neurone van PD patiënte te voorkom, te stop of om dit om te keer.

# Acknowledgements

First and foremost I would like to thank my supervisor Professor Soraya Bardien, for her constant support and guidance throughout my thesis and post-graduate studies. Soraya I have great respect for you and appreciate all the hard work in making this project possible. You always go above and beyond for your students and for that I am truly grateful.

To the Parkinson's disease group, past and present: Annika Neethling, Brigitte Glanzmann, Celia van der Merwe, Sihaam Boolay and William Haylett. Thank you for taking the time to always help and give advice.

I would like to thank Prof Sian Hemmings for all the help and advice during this research project.

Thanks are also due to the Shared Roots team and the MAGIC lab for all the support. I would like to sincerely thank Prof. Paul van Helden for the opportunities you provided.

I would like to express my gratitude to, Professor Helena Kuivaniemi and Professor Gerard Tromp for taking the time to help me with a large part of this research and always motivating me.

Alma Polson, Nicholas Bowker Lyndon Zass and Ruschca Jacobs thank you for all laughs and much needed coffee breaks.

A special thank you to my family and friends who motivated, supported and believed in me throughout this process. To my parents, I am forever grateful for all the love, encouragement and words of wisdom. To my sister, Bianca thank you for all the motivation and love.

I am thankful to God for giving me faith, strength, hope and direction.

I would like to acknowledge the funding bodies, the National Research Foundation, the Harry Crossley Foundation. Stellenbosch University and the Department of Biomedical Sciences for use of their facilities and the opportunity to further my academic career.

This study would not have been possible without the study participants and for this I would like to thank you. I hope that through our research we will one day be able to cure Parkinson's disease.

# Table of Contents

Declaration ..... i
Abstract ..... ii
Opsomming ..... iv
Acknowledgements ..... vi
List of Figures ..... ix
List of Tables ..... xi
List of abbreviations ..... xiii
Chapter 1: Introduction and Literature Review ..... 1
1.1 Introduction to Parkinson's disease (PD) ..... 2
1.2 Pathology ..... 3
1.3 Prevalence and Incidence ..... 5
1.4 Genetic Etiology ..... 6
1.5 Pathways and Disease Mechanisms ..... 8
1.6 Transcriptomics to Identify New Pathways in PD ..... 18
1.7 The Present Study ..... 38
Chapter 2: Materials and Methods ..... 41
2.1 Overview of the Methodology ..... 43
2.2 Study Participants ..... 44
2.3 Ethical considerations ..... 46
2.4 DNA Isolation and Methods ..... 46
2.5. RNA Isolation and Sequencing ..... 54
2.6 Bioinformatic Data Analysis ..... 58
2.7 Gene Prioritisation ..... 62
2.8 Verification using qPCR ..... 63
Chapter 3: Results ..... 65
3.1 Screening for pathogenic mutations ..... 68
3.2 RNA Isolation and Sequencing ..... 73
3.3 RNA-Seq Analysis Pipeline ..... 82
3.4 Enrichment Analysis ..... 92
3.5 Gene Prioritization ..... 98
3.6 Verification of *CEPBA* gene ..... 102

Chapter 4: Discussion ..... 103
4.1 Detection of a heterozygous *PARK2* exon 2 deletion ..... 107
4.2 Quality Control of RNA and cDNA ..... 109
4.3 RNA-Seq Pipeline ..... 110
4.4 Biological pathways and enrichment analysis ..... 112
4.5 Link between *CEBPA*, *PGC1α* and PD ..... 113
4.6 Verification using qPCR ..... 118
4.7 Study limitations ..... 119
4.8 Future Work ..... 120
4.9 Conclusions ..... 122
Reference List ..... 123
Appendix I ..... 141
Appendix II ..... 141
Appendix III ..... 142
Appendix IV ..... 143
Appendix V ..... 158
Appendix VI ..... 159
Appendix VII ..... 168
Appendix VIII ..... 169
Appendix IX ..... 170
Appendix X ..... 171
Appendix XI ..... 172
Appendix XII ..... 173
Appendix XIII ..... 174
Appendix XIV ..... 178
Appendix XV ..... 179
Appendix XVI ..... 180
Appendix XVII ..... 181

# List of Figures


| | | Page |
|---|---|---|
| **Chapter 1** | | |
| Figure 1.1 | A representation of the brain regions affected by Parkinson's disease | 3 |
| Figure 1.2 | A representation of the pathology of Parkinson's disease | 4 |
| Figure 1.3 | Braak staging of Parkinson's disease and symptoms associated with each stage | 5 |
| Figure 1.4 | Biological pathways and mechanisms involved in Parkinson's disease pathogenesis | 9 |
| Figure 1.5 | The ubiquitin proteasome system and it's involvement in Parkinson's disease | 11 |
| Figure 1.6 | RNA-Seq analysis workflow | 20 |
| **Chapter 2** | | |
| Figure 2.1 | Overview of methodology used for this study | 43 |
| Figure 2.2 | Multiplex ligation-dependent probe amplification (MLPA) assay principle and the various steps involved in the experiment | 49 |
| Figure 2.3 | PCR products of the *HBB* primer set | 54 |
| Figure 2.4 | Bioinformatic analysis pipeline of RNA-Seq data using Partek® Flow® | 59 |
| **Chapter 3** | | |
| Figure 3.1 | MLPA analysis results from a male PD patient SR023 with a heterozygous deletion of PARK2 exon 2 (ratio<0.6) | 70 |
| Figure 3.2 | Quantitative real-time PCR verified a *PARK2* exon 2 heterozygous deletion in patient SR023 | 71 |
| Figure 3.3 | Sanger sequencing of *PARK2* exon 2 | 72 |
| Figure 3.4 | RNA quality and concentration as assessed by the Agilent 2100 Bioanalyser | 75-76 |
| Figure 3.5 | Amplification of cDNA and gDNA templates using the *HBB* primer set | 77 |
| Figure 3.6 | Flowchart of the RNA-Seq data analysis pipeline | 83 |
| Figure 3.7 | Cluster heatmap illustrating differential gene expression between PD patient and control samples | 91 |
| Figure 3.8 | The five canonical pathways found to be dysregulated between Parkinson's disease patients and controls using Ingenuity Pathway Analysis | 94 |
| Figure 3.9 | Gene expression network from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using Ingenuity Pathway Analysis | 95 |
| Figure 3.10 | Gene expression network from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using Ingenuity Pathway Analysis | 96 |
| Figure 3.11 | Gene expression network from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using Ingenuity Pathway Analysis | 97 |
| Figure 3.12 | Clustering of cases and controls according to reads using Partek Flow for each prioritized candidate gene. | 99-100 |
| Figure 3.13 | Verification of differential gene expression of *CEBPA* by qPCR using the | 104 |

Lightcycler® 96
Figure 3.14 Verification of differential gene expression of *CEBPA* by qPCR 105

## Chapter 4

Figure 4.1 Proposed model of the interaction between *CEBPA* and *PGC-1α* and the link between Parkinson's Disease and Diabetes mellitus 117

# List of Tables


| | | Page |
|---|---|---|
| **Chapter 1** | | |
| Table 1.1 | Genes associated with Parkinson's disease and their known or putative function in the cell | 7 |
| Table 1.2 | Genes associated with Parkinson's disease which have been implicated in mitochondrial dysfunction | 17 |
| Table 1.3 | A summary of genome-wide mRNA and miRNA expression studies performed on blood and brain tissue from patients with Parkinson's disease | 22-33 |
| **Chapter 2** | | |
| Table 2.1 | Demographic and clinical characteristics of the 40 study participants | 44-45 |
| Table 2.2 | Comparison of demographic information of the all study participants | 45 |
| Table 2.3 | MLPA Master-mix reagents and volumes used for an MLPA assay | 50 |
| Table 2.4 | Copy number and dosage quotient values using Coffalyser.net software | 50 |
| Table 2.5 | Primers designed for the exonic sequences of the *PARK2* gene | 52 |
| Table 2.6 | Primer sequences for *HBB* reference gene | 53 |
| Table 2.7 | Partek E/M default input parameters | 60 |
| Table 2.8 | GSA input parameters | 60 |
| Table 2.9 | Filtered gene list parameters | 61 |
| Table 2.10 | Real time PCR primers and conditions | 63 |
| Table 2.11 | qPCR reaction master mix and set-up | 65 |
| Table 2.12 | qPCR cycling conditions using the Lightcycler®96 | 65 |
| **Chapter 3** | | |
| Table 3.1 | RNA concentration and purity for patient and control blood samples collected in PAXgene tubes using the Agilent 2100 Bioanalyser | 73-74 |
| Table 3.2 | RNA concentration, purity using the Agilent 2100 Bioanalyser. Comparison of the RIN values between internal (CAF) and external (Nxt-Dx) quality control | 78 |
| Table 3.3 | cDNA concentration and quality during cDNA library preparation | 80-82 |
| Table 3.4 | Total number of reads and read mapping for patient and control samples | 84 |
| Table 3.5 | Candidate gene list of differentially expressed genes in PD patients compared to controls using specified filtering parameters | 85 |
| Table 3.6 | Functional annotation clustering of 132 candidate genes using DAVID Bioinformatics database | 92 |
| Table 3.7 | Five significant canonical pathways found in whole blood from PD patients using Ingenuity pathway analysis | 93 |
| Table 3.8 | Five significant tox lists found in whole blood from PD patients using Ingenuity Pathway Analysis | 97 |
| Table 3.9 | Genes prioritized from candidate gene list based on a literature search | 97 |

| | | |
|---|---|---|
| | identifying genes in the same direction (up/ down) of differential expression | |
| Table 3.10 | Function of four prioritized candidate genes | 101 |
| Table 3.11 | Comparison of *CEBPA* expression to three reference genes *HBB*, *GAPDH* and *ACTB* using REST-2009© software | 102 |
| Table 3.12 | Non-normalized results of *CEBPA*, *HBB*, *GAPDH* and *ACTB* expression using REST-2009 © software | 103 |
| Table 3.13 | Comparison of *CEBPA* expression to two reference genes *HBB*, and *ACTB* using REST-2009 © software | 103 |

# List of abbreviations

| | |
|---|---|
| AAO | Age-at-onset |
| ACTB | Actin beta |
| AD | Autosomal dominant |
| AR | Autosomal recessive |
| ATP | Adenosine triphosphate |
| ATP13A2 | ATPase type 13A2 |
| BA9 | Brodmann's area 9 |
| BLAST | Basic local alignment search tool |
| bp | Base pairs |
| BR1 | Resuspension buffer |
| BR3 | Wash buffer 1 |
| BR4 | Wash buffer 2 |
| BR5 | Elution buffer |
| CAF | Central Analytical Facility |
| CAT | Catalase |
| CDCrel | Cell division control-related protein |
| cDNA | Complementary DNA |
| CE | Cerebellum |
| CEBPA | CCAAT/enhancer binding protein alpha |
| CHCHD2 | Coiled-coil-helix-coiled-coil-helix domain-containing protein 2 |
| CI | Complex I |
| *cIAP1* | Baculoviral IAP repeat-containing protein 2 |
| CNS | Central nervous system |
| CNV | Copy number variation |
| Cq | Quantification cycle |
| CVD | Cardiovascular disease |
| Cy3 | Cyanine 3 |
| Cy5 | Cyanine 5 |
| DA | Dopaminergic |
| DAT | Dopamine transporter |

| | |
|---|---|
| DAVID | Database for annotation, visualization and integrated discovery |
| D-fragments | Denaturation controls |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleotide |
| DRB | Dnase resuspension buffer |
| E/M | Expectation-maximization |
| E1 | Ubiquitin activating enzyme |
| E2 | Ubiquitin conjugating enzyme |
| E3 | Ubiquitin ligating enzyme |
| EDTA | Ethylene-diamine-tetra-acetic acid |
| EIF2 | Eukaryotic initiation factor 2 |
| eIF4 | Eukaryotic initiation factor 4 |
| EIF4G1 | Eukaryotic translation initiation factor 4 gamma 1 |
| ER | Endoplasmic reticulum |
| EtBr | Ethidium bromide |
| ETC | Electron transport chain |
| Exo I | Exonuclease |
| FBP1 | Far upstream sequence element binding protein 1 |
| FBXO7 | F-box protein 7 |
| FDR | False discovery rate |
| Fe-S | Iron-sulphur |
| GAPDH | Glyceraldehyde-3-phosphate dehydrogenase |
| gDNA | Genomic DNA |
| GEO | Gene expression omnibus |
| GSA | Gene specific analysis |
| GSH | Glutathione |
| Ha | Alternative Hypothesis |
| HAMD-D | Hamilton depression scale |
| HBB | Haemoglobin beta |
| HFHSD | High fat, high sucrose diet |
| HLOE | Highest level of education |

| | |
|---|---|
| HNF4A | Hepatocyte nuclear factor |
| Ho | Null Hypothesis |
| HREC | Health research ethics committee |
| Hsc70 | Heat-shock cognate protein |
| Hsp70 | Heat-shock protein 70 |
| HSP90AA1 | Heat shock protein 90 kDa alpha |
| HSPA8 | Heat shock 70 kDa protein 8 |
| HTRA2 | HtrA serine peptidase 2 |
| IFN-γ | Interferon-γ |
| IL-1 | Interleukin 1 |
| IPA | Ingenuity pathway analysis |
| KEGG | Kyoto encyclopedia of genes and genomes |
| LB | Lewy body |
| LN | Lewy neurite |
| lncRNA | Long-non-coding RNA |
| LPS | Lipopolysaccharide |
| LRRK2 | Leucine-rich repeat kinase 2 |
| M | Million |
| MAO | Monoamine oxidase |
| MAPK | Mitogen-activated protein kinases |
| MCT | Microcentrifuge tube |
| MetS | Metabolic syndrome |
| MLPA | Multiplex ligation-dependent probe amplification |
| Mn-EBDC | Manganese ethylene-bis-dithiocarbamate |
| MODY | Maturity onset diabetes of young |
| MPTP | 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine |
| mRNA | Messenger RNA |
| NCBI | National center for biotechnology information |
| NF-κB | Nuclear factor-κB |
| $NH_4$ | Ammonium |
| NPDs | Neuropsychiatric disorders |
| OMIM | Online mendelian inheritance in man |

| | |
|---|---|
| Pael-R | Parkin-associated endothelin-like receptor |
| PARK2 | Parkin |
| PBMC | Peripheral blood mononuclear cell |
| PCR | Polymerase chain reaction |
| PD | Parkinson's disease |
| PE | Paired-end |
| PGC-1α | Peroxisome proliferator-activated receptor gamma coactivator-1α |
| PINK1 | PTEN-induced putative kinase 1 |
| PK | Proteinase K |
| PMI | Post-mortem delay |
| PPAR | Peroxisome proliferator activated receptor |
| PRC | PAXgene RNA spin column |
| PSC | PAXgene shredder spin column |
| PT | Processing tube |
| PTBP1 | Polypyrimidine tract binding protein 1 |
| PUFA | Polyunsaturated free fatty acids |
| PUT | Putamen |
| QC | Quality control |
| Q-fragments | Quality control probes |
| qPCR | Quantitative real-time PCR |
| RDD | DNA digestion buffer |
| REDCap | Research electronic data capture |
| RIN | RNA intergrity number |
| RNA | Ribonucleic acid |
| RNA-Seq | RNA sequencing |
| ROS | Reactive oxygen species |
| rRNA | Ribosomal RNA |
| -RT | Non-reverse transcription control |
| RXR | Retinoid X receptors |
| Sap I | Shrimp alkaline phosphatase |
| SB | Sodium borate |
| SCF | Skp-Cullin-F-Box |

| | |
|---|---|
| SE | Single-end |
| SN | Substantia nigra |
| SNCA | α-synuclein |
| SNP | Single nucleotide polymorphism |
| SNpc | Substantia nigra pars compacta |
| SOD | Superoxide dismutase |
| ST13 | Suppression of tumorigenicity 13 |
| STAR | Spliced transcripts alignment to a reference |
| SYNJ1 | Synaptojanin 1 |
| TMEM20 | Transmembrane protein 20 |
| TNF | Tumour necrosis factor |
| TXLNG2P | Taxilin gamma pseudogene, Y-linked |
| UPDRS | Unified Parkinson's disease rating scale |
| UPS | Ubiquitin-proteasome system |
| USP9Y | Ubiquitin specific peptidase 9, Y-linked |
| UTY | Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked |
| VPS35 | Vacuolar protein sorting 35 |
| WHO | World Health Organization |
| WT | Wild-type |

# Chapter 1: Introduction and Literature Review

1.1 Introduction to Parkinson's disease (PD) ..... 2
1.2 Pathology ..... 3
1.3 Prevalence and Incidence ..... 5
1.4 Genetic Etiology ..... 6
1.5 Pathways and Disease Mechanisms ..... 8
1.5.1 The Ubiquitin-Proteasome System ..... 10
1.5.2 Oxidative Stress ..... 12
1.5.3 Mitochondrial Dysfunction ..... 14
1.6 Transcriptomics to Identify New Pathways in PD ..... 18
1.6.1 High-throughput Technologies for Transcriptomic Profiling ..... 18
1.6.2 Summary of Previous Transcriptomic Studies in PD ..... 21
1.7 The Present Study ..... 38
1.7.1 Research Problem Statement ..... 38
1.7.2 Hypotheses ..... 40
1.7.3 Research Aim and Objectives ..... 40

## 1.1 Introduction to Parkinson's disease (PD)

Neurological disorders are one of the leading causes of death worldwide and constitute 12% of the total deaths globally, according to a report published by the World Health Organization (WHO) in 2016 (WHO, 2016). Worldwide, populations are undergoing a demographic transition with people becoming progressively older and therefore, neurological disorders affecting older individuals are set to become significantly more common. One such neurological disorder is Parkinson's disease (PD) which is the focus of the current study.

PD was first documented in 1817 by English physician Dr. James Parkinson where he reported on this disorder in an essay entitled "An essay on the shaking palsy" in which three clinical motor symptoms were identified including decreased muscle strength, loss of posture and gait, and resting tremor (Parkinson, 1817). Dr Jean Martin Charcot formally recognized PD approximately 60 years after this first documentation and named the disease in honour of Dr James Parkinson.

PD is currently the most common movement disorder and the second most common neurodegenerative disease after Alzheimer's disease. Pathological hallmarks of idiopathic PD include the presence of α-synuclein inclusions in Lewy bodies (LBs) (Spillantini et al., 1997) and loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc) (Braak et al., 2003) which is associated with the nigrostriatal pathway (Dauer and Przedborski, 2003). Reduction in levels of dopamine leads to dysregulation of motor circuits (Figure 1.1). Pathological degeneration of dopaminergic neurons results in the four cardinal motor-associated symptoms including resting tremor (shaking of the hands and fingers even when at rest), rigidity (stiff or inflexible muscles), bradykinesia (slowness of movement) and postural instability (loss of stability and balance) (Xia and Mao, 2012). A number of non-motor symptoms are also associated with the disease. These include loss of smell (Hawkes et al., 1999) and psychiatric symptoms such as depression, anxiety, and dementia (Hu et al., 2011).

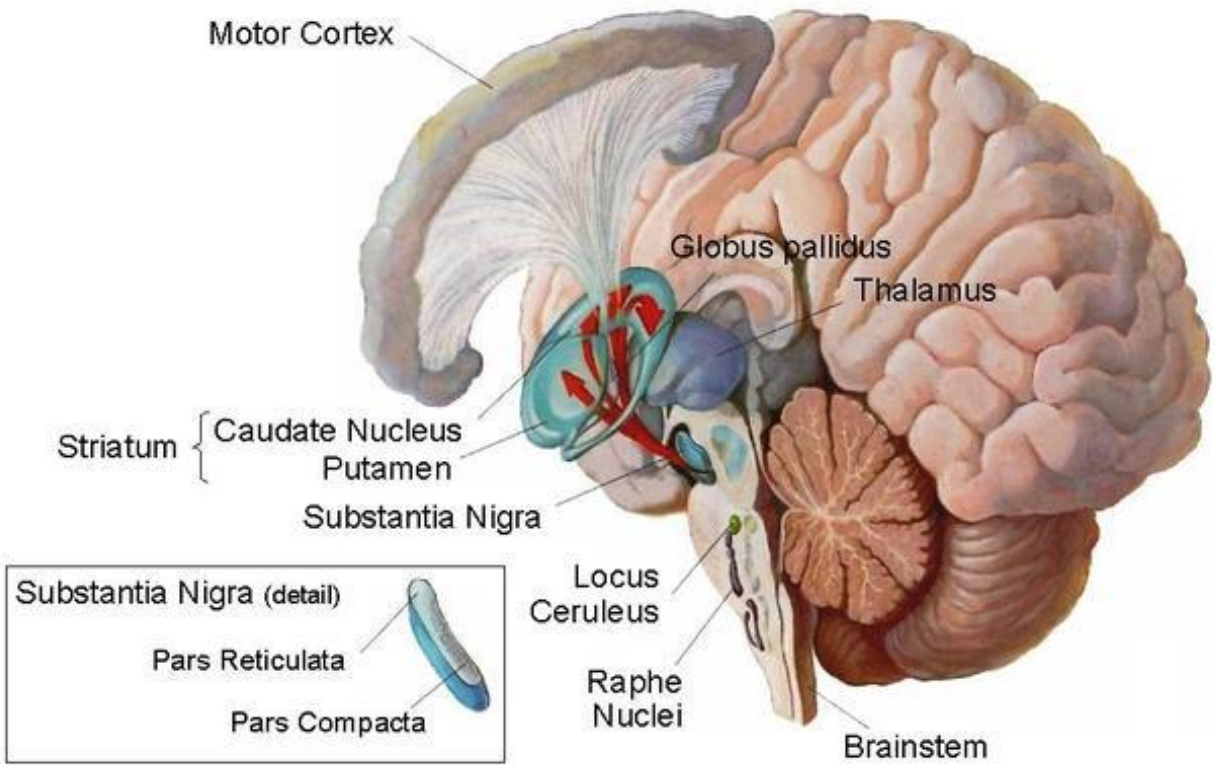



**Figure 1.1: A representation of the brain regions affected by Parkinson's disease.** These regions form part of the basal ganglia and nigrostriatal pathway. Taken from https://stepsys.files.wordpress.com/2013/11/parkinsons-disease-204114724_std.jpg

## 1.2 Pathology

PD pathology is relatively well understood and primarily characterized by loss of dopaminergic neurons and reactive gliosis in the SNpc (Dickson, 2012). PD can only be diagnosed with certainty by autopsy. Anatomical cut sections of the PD brain show a loss of dark pigment in the substantia nigra (SN) and locus ceruleus (Figure 1.2). Depigmentation of these regions correlates with dopaminergic neuronal loss, as these neurons contain neuromelanin which is responsible for the dark color of the SN. At autopsy, most cases show a 70-80% dopaminergic neuronal loss (Figure 1.2B) whereas in unaffected individuals the SN is highly pigmented (Figure 1.2A).

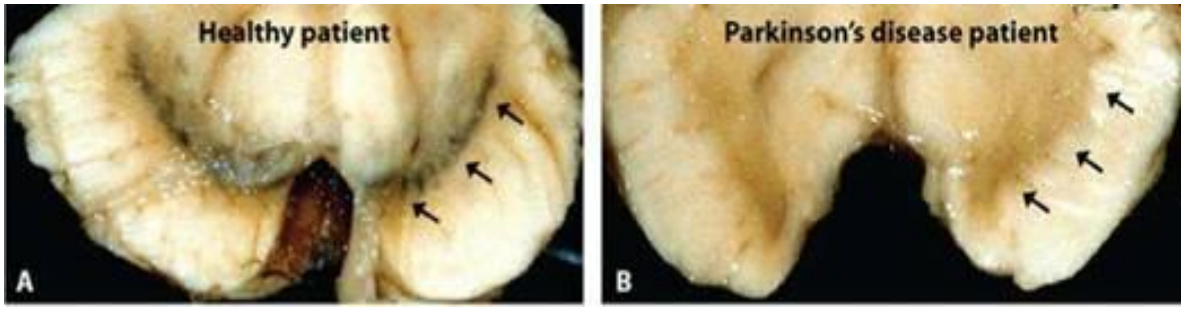



**Figure 1.2: A representation of the pathology of Parkinson's disease (PD)**. Sections of the midbrain from (A) healthy patient and (B) PD patient. The black arrows indicate the substantia nigra showing de-pigmentation in the affected individual as a result of loss of neuromelanin in dopaminergic neurons. (Taken from http://10sa.org/wp-content/uploads/114/image005_34.jpg)

A second pathological hallmark of idiopathic PD is the presence of inclusion bodies known as LBs and amyloid fibrils known as Lewy neurites (LNs) (Volpicelli-Daley et al., 2014). The majority of PD patients present with these inclusions which occur in the SN and several other brain regions. LBs are intracytoplasmic neuronal inclusions which are composed of neurofilaments and ubiquitin. However, aggregated and misfolded α-synuclein protein has been identified as a major component of LBs (Spillantini et al., 1997). Therefore, idiopathic PD is defined as a synucleinopathy, due to this excessive accumulation of α-synuclein in neurons. Notably, the Braak Hypothesis states that PD pathology is a progressive process which occurs over various stages and over many years. It is hypothesized that the formation of LBs and LNs begins at a defined site and progresses sequentially (Figure 1.3; Braak et al., 2003). The inclusion body pathology occurs initially in the medulla oblongata, pontine tegmentum, olfactory bulb and anterior olfactory nucleus. Thereafter, this progresses to the SN and finally reaches the neocortex (Braak et al., 2004). Therefore, non-motor associated symptoms such as the loss of smell is thought to be indicative of a presymptomatic stage of the disease (Braak stages 1 and 2 which occurs up to 10 years prior to PD diagnosis) and the olfactory bulb is thought to be one of the earliest regions affected by LB formation (Braak et al., 2004).

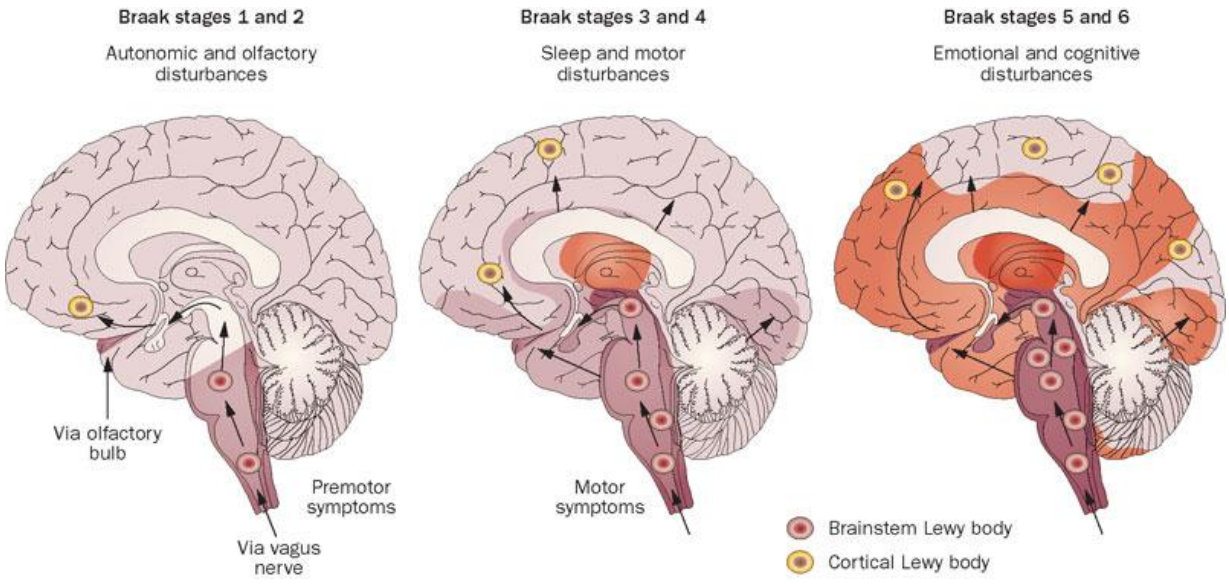



**Figure 1.3: Braak staging of Parkinson's disease and symptoms associated with each stage.** Stages 1 and 2 represent the presymptomatic stages of the disease where inclusion body pathology is confined to the medulla oblongata and olfactory bulb. Stages 3 and 4 include motor associated symptoms as the subtantia nigra and other regions of the midbrain are affected. Severe pathological changes result in stages 5 and 6 where the pathological process occurs in the neocortex. Taken from Doty, 2012

## 1.3 Prevalence and Incidence

The prevalence of a disease is a measurement of the number of cases of that disease that are present in a particular population at a given time (Sellbach et al., 2006). PD currently affects 1% of individuals over the age of 60, with a global prevalence of 3% (de Lau and Breteler, 2006). By the year 2030, it is projected that 9 million individuals will be affected by PD therefore posing a major global burden. The estimated prevalence of PD in developed countries is set to increase by 92% between 2010 and 2050 (Bach et al., 2011). Notably, the speed of population ageing in sub-Saharan Africa will exceed that of industrialized countries making diseases associated with ageing, an even greater disease burden for these countries. This is borne out by data from Tanzania which shows that the prevalence of PD in that country is predicted to increase by an alarming 184% by 2025 (Dotchin et al., 2012)

Incidence refers to the rate at which individuals with a particular disorder are newly diagnosed (Eeden et al., 2003). A recent systematic review and meta-analysis investigated the incidence of PD and it's variation by age and gender (Hirsch et al., 2016). It was found there was a higher incidence rate in males compared to females between the ages of 60-69 years. The incidence rate in males was 58.22 per 100

000 individuals whereas in females it was 30.32 per 100 000 individuals. The study also identified that PD incidence increases with age.

PD is mainly regarded as a sporadic disease generally occurring after the age of 60 years and this is referred to as idiopathic PD. Up to 10% of PD cases are, however, familial and thought to have a genetic basis (Klein and Westenberger, 2012).

## 1.4 Genetic Etiology

PD is thought to arise due to a combination of environmental and genetic factors. Environmental factors include pesticides such as paraquat (Betarbet et al., 2000), and mitochondrial poisons such as 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (Sian et al., 1999) and rotenone (Sherer et al., 2003b). To date, mutations in 11 genes have been linked to heritable monogenic forms of PD (Table 1.1) which have been verified in different populations. Mutations in α-synuclein (*SNCA*) (Polymeropoulos et al., 1997), leucine-rich repeat kinase 2 (*LRRK2)* (Paisán-Ruíz et al., 2004; Zimprich et al., 2004), vacuolar protein sorting 35 (*VPS35*) (Vilarino-Guell et al., 2011; Zimprich et al., 2011), coiled-coil-helix-coiled-coil-helix domain-containing protein 2 (*CHCHD2*) (Funayama et al., 2015) and transmembrane protein 20 (*TMEM20*) (Deng et al., 2016) lead to autosomal dominant (AD) PD; whereas autosomal recessive (AR) inheritance of PD is caused by mutations in *Parkin (PARK2)* (Kitada et al., 1998), PTEN-induced putative kinase 1 (*PINK1)* (Valente et al., 2004), *DJ-1* (Bonifati et al., 2003), ATPase type 13A2 (*ATP13A2)* (Ramirez et al., 2006) F-box protein 7 (*FBX07*) (Shojaee et al., 2008) and synaptojanin 1 (*SYNJ1*) (Krebs et al., 2013; Quadri et al., 2013). Eukaryotic translation initiation factor 4 gamma 1 (*EIF4G1)* (Chartier-Harlin et al., 2011) has also been implicated in PD, which causes AD PD however, further validation is required in other populations.

There is also increasing evidence that other genetic risk factors increase PD susceptibility which include a functional repeat polymorphism in *REP1*, the *SNCA* promoter (Mata et al., 2010) and increased phosphorylation of TAU protein (Reviewed by: Lei et al., 2010).

**Table 1.1: Genes associated with Parkinson's disease and their known or putative function in the cell.**

| Locus | Gene | Inheritance | Function |
|---|---|---|---|
| PARK1 | *SNCA* | AD | • Regulation of dopamine release and transport (Subramaniam et al., 2014) |
| PARK2 | *Parkin* | AR | • E3 ubiquitin ligase which functions in the Ubiquitin proteasome system (Shimura et al., 2000)<br>• Maintenance of mitochondria (Narendra et al., 2008)<br>• Removal of damaged mitochondria via mitophagy (Narendra et al., 2010; Geisler et al., 2010)<br>• Involved in the UPS (Chan and Chan, 2011) |
| PARK6 | *PINK1* | AR | • Maintenance of mitochondria by protecting against apoptosis (Petit et al., 2005; Plun-Favreau and Hardy, 2008)<br>• Removal of damaged mitochondria (Narendra et al., 2010; Geisler et al., 2010)<br>• Regulation of mitochondrial fission machinery (Hoppins et al., 2007) |
| PARK7 | *DJ-1* | AR | • Involved in the antioxidative stress reaction (Yokota et al., 2003; Taira et al., 2004; Yanagida et al., 2009)<br>• Present in the mitochondria and protects against oxidative neuronal death (Canet-Avilés et al., 2004) |
| PARK8 | *LRRK2* | AD | • Functions in the immune system by activating signaling pathways linked to immune response (Gardet et al., 2010)<br>• Localised to the outer mitochondrial membrane (Biskup et al., 2006; West et al., 2005) |
| PARK9 | *ATP13A2* | AR | • Regulates lysosomal functions (Ramirez et al., 2006; Tsunemi and Krainc, 2014) |
| PARK15 | *FBX07* | AR | • Constitutes the ubiquitin protein ligase complex, functions in phosphorylation–dependent ubiquitination (Nelson et al., 2013)<br>• Regulation of hematopoiesis (van der Harst et al., 2012) |
| PARK17 | *VPS35* | AD | • Retrograde transport of proteins from endosomes to the trans-Golgi network (Bonifacino and Rojas, 2006) |
| Not assigned | *CHCHD2* | AD | • Functions in dopamine transporter binding in the putamen and caudate nucleus (Shi et al., 2016) |
| Not assigned | *TMEM20* | AD | • Encodes a transmembrane protein of secretory and recycling vesicles (Deng et al., 2016)<br>• Involved in synaptic vesicle trafficking (Deng et al., 2016) |
| Not assigned | *SYNJ1* | AR | • Endocytic traffic regulation at synapses (Krebs et al., 2013)<br>• Synaptic vesicle recycling and modulation of synaptic activity (Drouet et al., 2014) |

AD, Autosomal dominant; AR, Autosomal recessive; *CHCHD2,* Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 2; *FBXO7*, F-Box Protein 7; *LRRK2*, Leucine-rich repeat kinase 2*; PINK1*, PTEN-induced putative kinase 1; *SNCA*, α-synuclein; *SYNJ1,* Synaptojanin 1; *TMEM20,* Transmembrane protein 20; UPS, Ubiquitin proteasome system; *VPS35,* Vacuolar protein sorting-associated protein 35

## 1.5 Pathways and Disease Mechanisms

Identification of genes associated with PD has aided in our understanding of the biological pathways implicated in the disease. Several pathways and processes identified thus far include the ubiquitin-proteasome system (UPS) (Martins-Branco et al., 2012), mitophagy of dysfunctional mitochondria (Narendra et al., 2009), oxidative phosphorylation (Liu et al., 2011), neuroinflammation (reviewed by: Wang et al., 2015), autophagy-lysosome pathway (reviewed by: Gan-Or et al., 2015) and apoptosis (reviewed by: Lev et al., 2003). Possible mechanisms associated with PD pathogenesis have also been investigated including oxidative stress and mitochondrial dysfunction. Figure 1.4 represents an overview of the biological pathways involved in PD pathogenesis and some of the main ones shall be elaborated on in this section.

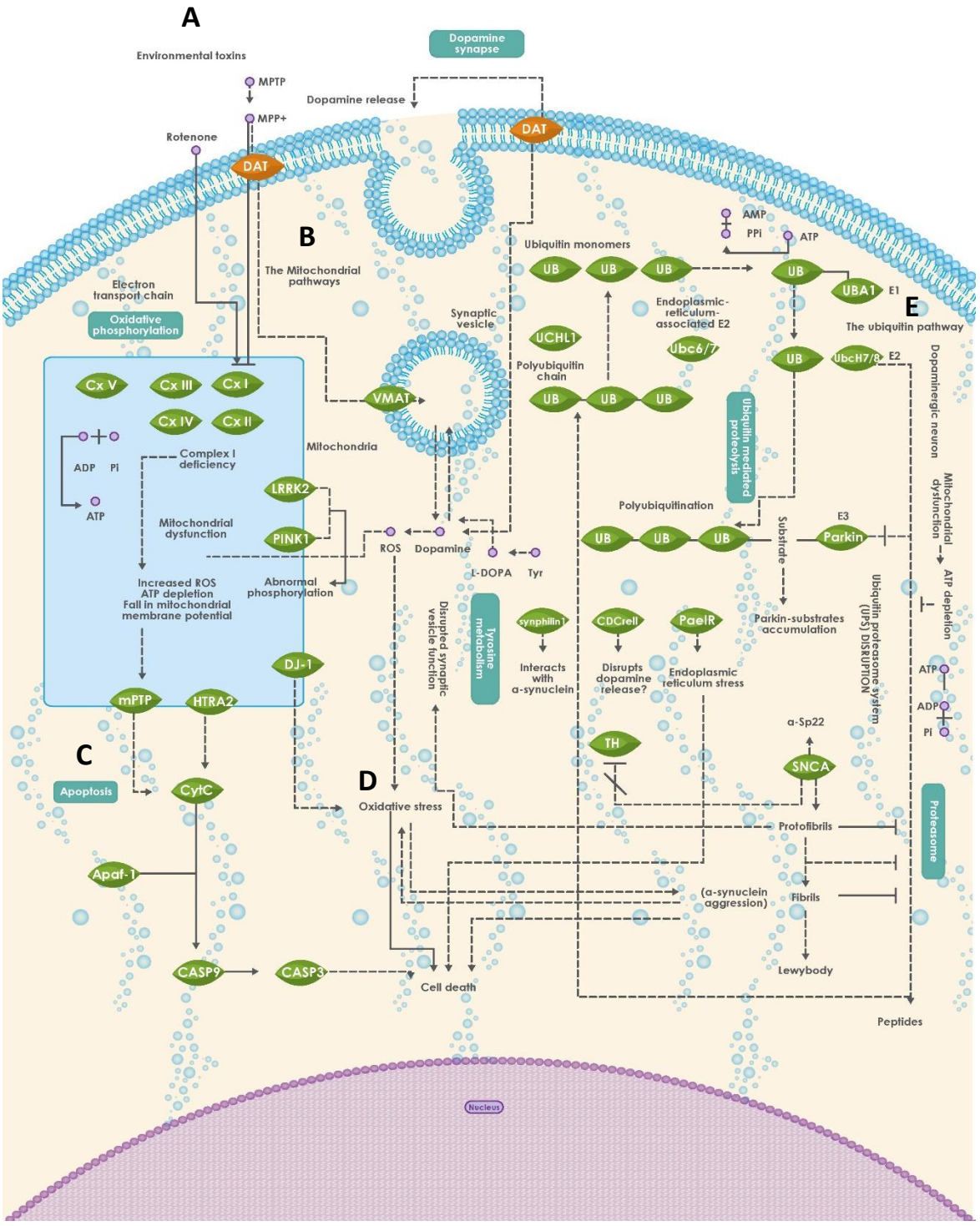



**Figure 1.4: Biological pathways and mechanisms involved in Parkinson's disease pathogenesis.** A. Environmental toxins (MPTP and Rotenone) resulting in cell death. B. The mitochondrial pathways and mitochondrial dysfunction. C. Apoptotic pathway and cause of cell death. D. Oxidative stress. E. The ubiquitin proteasome system. Taken from Aviva Systems Biology, 2013 (http://www.avivasysbio.com)

### 1.5.1 The Ubiquitin-Proteasome System

The ubiquitin-proteasome system (UPS) mediates the degradation of abnormal, unwanted or misfolded proteins in the cytoplasm, nucleus and endoplasmic reticulum (ER) of cells (Hershko and Ciechanover, 1998) thereby maintaining protein homeostasis. The UPS functions sequentially by tagging proteins destined for degradation with a 76 amino acid residue protein named ubiquitin. The C-terminus of ubiquitin is covalently attached to lysine side chains of the target protein via an iso-peptide linkage, in a process known as ubiquitination. The carboxyl end of the ubiquitin is activated in an ATP-dependent process by the ubiquitin activating enzyme (E1), the activated ubiquitin which has a highly reactive ubiquitin thiolester binds to the ubiquitin conjugating enzyme (E2) which catalyses the transfer of the ubiquitin onto the E2 active site. The E2 enzyme then transfers ubiquitin to the ubiquitin ligating (E3) enzyme. The E3 ligase allows the transfer of ubiquitin from the E2 enzyme to the protein substrate. The ubiquination process is repeated several times to allow for the formation of a polyubiquitin chain (Figure 1.5, labelled A). Iso-peptidic bonds between the G76 and K48 residues link the ubiquitin molecules together (Chau et al., 1989) resulting in a G76-K48 polyubiquinated substrate which is targeted for degradation by the 26S proteasome . The proteasome binds and removes the polyubiquitin chain and unfolds the protein. The protein is subsequently threaded through the proteasome chamber where it is proteolysed into short peptides where it can be reused for the synthesis of new proteins and the ubiquitin is also recycled.

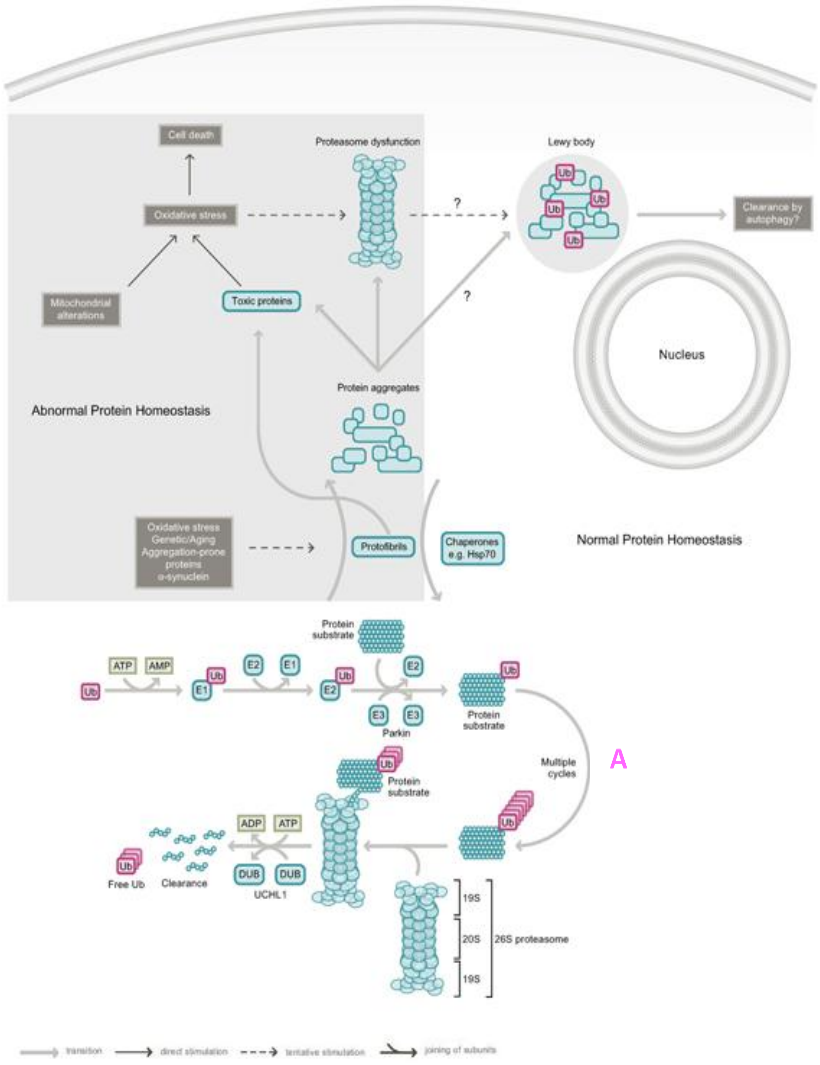



**Figure 1.5: The ubiquitin proteasome system and it's involvement in Parkinson's disease.** The components required for formation of a polyubiquitin chain for degradation by the 26S proteasome is shown (A). Taken from Lim et al, 2007.

Dysfunction in the UPS has been linked to PD pathogenesis (Leroy et al., 1998; McNaught and Jenner, 2001; McNaught et al., 2003). PD-causing mutations in the *PARK2* gene (which is a HECT/RING hybrid E3 ubiquitin ligase) provide evidence implicating the direct role of the UPS in PD (Imai et al., 2000; Shimura et al., 2000; Zhang et al., 2000). *PARK2* is involved in the ubiquination pathway for removal of misfolded proteins in the ER and therefore is thought to be involved in protection from neurotoxicity. *PARK2* could also possibly influence protein aggregation through regulation of molecular chaperone heat-shock protein 70 (Hsp70) (Moore et al., 2008). Disruption in protein homeostasis results in accumulation of intracellular proteins which are toxic and could cause neuronal loss. Accumulation of certain PARK2

substrates including cell division control-related protein (CDCrel) 1 and 2a (Choi et al., 2003), parkin-associated endothelin-like receptor (Pael-R) (Imai et al., 2001), cyclin E (Staropoli et al., 2003), p38 (Ko et al., 2005) and far upstream sequence element binding protein 1 (FBP1) (Ko et al., 2006) could induce neurotoxicity by disrupting cellular functions (reviewed in Sandebring et al, 2012). Accumulation of CDCrel-1 and CDCrel-2a proteins have been found in human *PARK2* mutant brains (Choi et al., 2003), and the interaction between *PARK2* and CDCrel-1 could possibly disrupt dopamine release from cells (Dong et al., 2003). Whereas PaelR accumulation results in ER stress and unfolded protein-induced cell death (Imai et al., 2001). Therefore, loss of functional PARK2 results in the accumulation of toxic substrates which could lead to neurodegeneration. Synphilin-1 and some septins which are α-synuclein binding proteins are also found to be substrates of PARK2. It has been shown that expression of PARK2, α-synuclein and synphilin-1 results in Lewy-body like ubiquitin-positive cytosolic inclusions (Chung et al., 2001).

*FBX07* is also involved in the UPS where it has been found to be an adaptor protein in the Skp-Cullin-F-Box (SCF). $SCF^{FBXO7}$ ubiquitin E3 ligase complex recognizes substrates for ubiquitination by $SCF^{FBXO7}$ E3 ligase (Zhou et al., 2016). *FBXO7* promotes ubiquitination and degradation of baculoviral IAP repeat-containing protein 2 known as *cIAP1* (an inhibitor of apoptosis which regulates Nuclear factor-κB (NF-κB) signaling) and therefore plays a role in cell apoptosis and *cIAP1* function (Chang et al., 2006). PD-causing mutations in *FBXO7* are thought to result in *FBXO7* protein aggregation causing proteotoxic stress and ultimately neurodegeneration.

### 1.5.2 Oxidative Stress

Oxidative stress is thought to be a central mechanism that causes cellular dysfunction and and could be argued as a critical factor leading to cellular demise (reviewed by, Blesa et al., 2015). Proteins in the central nervous system (CNS) can also be damaged not only by dysfunction of protein degradation pathways but also by reactive oxygen species (ROS) and other free radicals. Molecular oxygen is required by cells for catalysis and energy production and ROS are produced as byproducts. ROS are important for normal biological functions however excessive production is detrimental to cell membranes and could result in cell death.

Oxidative stress occurs in cells when there is excessive production of ROS and imbalance of antioxidant defense mechanisms resulting in damage to cellular proteins, lipids and deoxyribonucleic acid (DNA). The brain is composed of phospholipids and polyunsaturated free fatty acids (PUFAs) in their

membranes both of which are highly susceptible to oxidative damage. These cells also have a low endogenous antioxidant enzymatic activity. Notably, the SN is deficient in protective mechanisms against ROS, superoxide radicals and hydrogen peroxide and therefore is highly susceptible to oxidative stress (Dickson, 2007).

Post-mortem studies of PD patients showed increased levels of oxidized lipids, proteins and DNA and decreased levels of reduced glutathione (GSH) which indicates that oxidative stress is an established event in PD (Mariani et al., 2005; Uversky, 2004). Markers of oxidative stress were also evident in PD postmortem tissues where malondialdehyde (a marker of lipid oxidation) was increased in the SN of PD patients and the concentration of PUFA in the SN of PD patients has been shown to be decreased (Dexter et al., 1989). Lipid peroxidation is a sensitive marker of oxidative stress (Nikolova and Mancheva, 2012). Oxidative stress has therefore been hypothesized to be associated with both the initiation and progression of PD (Zhou et al, 2008). The major sources of oxidative stress are thought to be ROS generated through increased dopaminergic metabolism, increased free iron levels, mitochondrial dysfunction and neuroinflammation (Hwang, 2013).

#### *1.5.2.1 Dopaminergic Metabolism*

Catecholamines and dopamine (DA) have been shown to be an important source of free radicals in the brain (Dias et al., 2013). DA is a neurotransmitter stored in synaptic vesicles where it is stable. However, excessive cytosolic DA is metabolized by monoamine oxidase (MAO) to produce hydrogen peroxide or by autoxidation via enzymes to form quinones specifically DA quinone. Proteins such as $\alpha$-synuclein, parkin, DJ-1 and UCH-L1 can be modified by DA quinone (Hwang, 2013). Modification of $\alpha$-synuclein results in its conversion to a cytotoxic protofibril form and utilizes chaperone mediated autophagy to inhibit degradation of other proteins.

Neuromelanin has been shown to worsen the neurodegenerative process by triggering neuroinflammation. Neuromelanin also catalyzes superoxide radical dismutation to hydrogen peroxide which is highly toxic to neuronal cells and can also bind free radicals. In PD patients there is an increase in DA metabolism which may be a compensatory mechanism as there is a DA shortage in the brain. Therefore, DA metabolism results in the generation of harmful ROS and endogenous toxins and could contribute to oxidative stress in PD. It should be noted that there is strong evidence for the link between oxidative stress and mitochondrial dysfunction and this shall be further elaborated on in Section 1.5.3.

## 1.5.3 Mitochondrial Dysfunction

### *1.5.3.1 Mitochondria and Mitochondrial Biogenesis*

Mitochondria are found in most eukaryotic cells and provide the cell with cellular energy in the form of adenosine triphosphate (ATP) generated through oxidative phosphorylation. Neurons require large amounts of ATP from aerobic respiration for optimal functioning. Beside ATP generation, mitochondria are involved in a number of cellular processes including the control of cell division and growth, regulation of cell death via apoptosis, calcium homeostasis, haem biosynthesis and the formation of iron-sulphur (Fe-S) clusters.

The formation of new mitochondria within the cell is a process referred to as mitochondrial biogenesis which is defined as the growth and division of pre-existing mitochondria (Jornayvaz and Shulman, 2010). Activation of mitochondrial biogenesis occurs in response to increased cellular energy demand, cellular stress or environmental stimuli resulting in an increase in ATP production (Alam and Rahman, 2014). Several transcriptional regulators within the cell modulate mitochondrial biogenesis among these is the peroxisome proliferator activated receptor (PPAR) family (Madrazo and Kelly, 2008). PPARs regulate mitochondrial biogenesis via an activator called peroxisome proliferator-activated receptor gamma coactivator-1$\alpha$ (*PGC-1*$\alpha$) (Wenz, 2009, p. -1; López-Lluch et al., 2008). *PGC-1*$\alpha$ is the master regulator of mitochondrial biogenesis and oxidative phosphorylation (Puigserver and Spiegelman, 2003; Zheng et al., 2010).

### *1.5.3.2 Mechanisms of Mitochondrial Dysfunction*

Mitochondrial dysfunction can occur as a result of several factors including environmental, various genetic factors and dysfunction of the UPS. A large body of evidence has implicated mitochondrial dysfunction as having a central role in neuronal cell death associated with PD (Basso et al., 2004; Keeney et al., 2006).

The electron transport chain (ETC) present in mitochondria is primarily responsible for ATP generation and is a major source of ROS in eukaryotic cells (Chance et al., 1979). In this system molecular oxygen produced is reduced to water by the chain complexes and a small amount of superoxide is then produced by complexes I and III. Superoxide produced within the mitochondria is then converted by the enzyme manganese superoxide dismutase to hydrogen peroxide. As previously mentioned ROS are normally necessary for signaling events however excessive ROS causes oxidative stress in the cell which results in damage to cellular macromolecules and can lead to cell death.

The first evidence for the involvement of mitochondrial dysfunction in PD was described when reduction in mitochondrial complex I (CI) activity was found in the SNpc of PD patients (Schapira et al., 1989) and therefore could be a possible mechanism in the pathogenesis of PD (Mizuno et al., 1989). CI deficiency has also been identified in the frontal cortex (Parker et al., 2008) as well as peripheral tissues such as platelet mitochondria (Haas et al., 1995) and skeletal muscle (Bindoff et al., 1991) of PD patients. However, skeletal muscle results have not been consistent. As CI deficiency is present in both the SN and platelets this may be indicative of a systemic defect which can either be genetic or environmentally caused.

It is hypothesized that nigrostriatal neurons are particularly sensitive to dysfunction of CI. An animal model was developed to investigate this, and rotenone was utilized for the study as it is an inhibitor of CI in the mitochondrial ETC (Greenamyre et al., 1999). Using the rotenone model on male Sprague-Dawley rats, they demonstrated that inhibition of CI results in damage to dopaminergic nerve terminals and nigral cell bodies .Rotenone has been shown to cause oxidative damage and dopaminergic neuronal loss specifically in the midbrain and olfactory bulb (Sherer et al., 2003). Therefore, nigrostriatal neurons that undergo death in PD have selective vulnerability to CI dysfunction and therefore a defect in this complex could result in PD pathogenesis.

Many studies have been conducted on PD using MPTP-treated primate models (Bezard et al., 1997). MPTP is a neural toxin causing mitochondrial dysfunction and is known to induce Parkinsonism in humans (Davis et al., 1979; Sian et al., 1999). In one study Cynomolgus monkeys were injected daily with MPTP (Bezard et al., 1997), and the drug was administered in such a way as to mimic the slow progression of PD and observe the effect on the motor symptoms. MPTP is converted to $MPP^+$ by MAO-B which is then transported by the DA transporter (DAT) (Sian et al., 1999) either into DA nerve terminals or into the mitochondria where it inhibits CI (Lin and Beal, 2006) causing increased ROS and decreased production of ATP (Fabre et al., 1999). CI deficiency results in an increase in the release of electrons from the transport chain into the mitochondrial matrix which then reacts with oxygen to form ROS such as hydroxyl radicals and nitric oxide. In one study on rat skeletal muscle mitochondria, inhibition of the ubiquinone binding site ( an electron acceptor) of CI results in an increase in electron radicals (Lambert and Brand, 2004). ROS act on signaling molecules by causing lipid peroxidation or promoting excitotoxicity resulting in protein modification and eventually cell death. MPTP has also been shown to cause inhibition of complexes-III and IV (Desai et al., 1996). Complex-III inhibition can also result from exposure to manganese ethylene-bis-dithiocarbamate (Mn-EBDC) an active element in

maneb, a pesticide, which is speculated to contribute to PD pathogenesis (Zhang et al., 2003). Paraquat, a herbicide which is also linked to development of PD, is toxic to mitochondria as it results in the production of superoxide and ROS (Cochemé and Murphy, 2008)

### *PD genes with known Mitochondrial Function*

Mutations in a number of PD-associated genes including: *α-synuclein*, *PARK2*, *PINK1*, *DJ-1*, *LRRK2* and HtrA serine peptidase 2 (*HTRA2)* (Table 1.2) have been shown to cause mitochondrial dysfunction. Recently *FBXO7* has been implicated in DA neuron degeneration by causing mitochondrial toxicity (Zhou et al., 2016). *FBXO7* has also been shown to have a direct interaction with *PARK2* and *PINK1* and therefore participate in mitophagy as well as modulate *PARK2* translocation to the mitochondria (Burchell et al., 2013).

**Table 1.2: Genes associated with Parkinson's disease which have been implicated in mitochondrial dysfunction.**

| Gene | Genetic link to Mitochondrial dysfunction | Reference |
|---|---|---|
| *α-synuclein* | • Mutant *α-synuclein* accumulates in inner mitochondrial membrane causing complex I impairment and increase in ROS | • Devi et al., 2008 |
| *PARK2* | • Interact with and promote degradation of *α-synuclein*<br>• *PARK2* null mice: reduction in subunits of Complex I and IV, reduced mitochondrial respiratory chain function, increased oxidative stress<br>• Reduced complex I activity in PD patients.<br>• DA neuronal degeneration and mitochondrial abnormalities in *PARK2* mutant drosophila. | • Kim et al., 2003<br>• Palacino et al., 2004<br>• Müftüoglu et al., 2004<br>• Wang et al., 2007 |
| *PINK1* | • Knockdown of *PINK1* induces mitochondrial oxidative stress, mitochondrial fragmentation, autophagy and dyregulation of calcium homeostasis in SH-SY5Y cells | • Gandhi et al., 2009; Dagda et al., 2009 |
| *DJ-1* | • Loss of *DJ-1* results in mitochondrial fragmentation, impaired dynamics, induced oxidative stress and autophagy | • Irrcher et al., 2010; Krebiehl et al., 2010 |
| *FBXO7* | • Protein dysfunction and deficient ubiquination of substrates<br>• Causes mitochondrial toxicity as mutations of *FBXO7* cause FBXO7 protein aggregation in the mitochondria | • Lohmann et al., 2015<br>• Zhou et al., 2016 |
| *HTRA2* | • *HTRA2* mutations cause increased ROS and defective mitochondrial respiration<br>• Deletion of *HTRA2* caused an accumulation of misfolded proteins in the mitochondria | • Moisoi et al., 2008 |
| *LRRK2* | • Drosophilia with *LRKK2* mutations increased susceptibility to complex I inhibitor and rotenone and increased protection against DA neurodegeneration | • Ng et al., 2009 |

*FBXO7*, F-box protein 7; *HTRA2*, HtrA serine peptidase 2; *LRRK2*, Leucine-rich repeat kinase 2; PD, Parkinson's disease; *PINK1*, PTEN-induced putative kinase 1; ROS, reactive oxygen species

In summary, although there is strong evidence for involvement of various biological processes in PD including mitochondrial dysfunction, the exact pathobiology and sequence of events that lead to development of PD is still unknonwn. The next section of this thesis will deal with transcriptomics which links the genome with the proteome and therefore could provide important clues to PD pathogenesis.

## 1.6 Transcriptomics to Identify New Pathways in PD

Transcriptomics is the study of the complete set of ribonucleic acid (RNA) transcripts expressed by an organism or in a particular cell or tissue type under certain conditions. This approach is a potential avenue for providing new insights into pathological events such as alternative splicing and differential gene expression, and has rapidly evolved and now involves high-throughput techniques such as microarrays and RNA sequencing (RNA-Seq).

### 1.6.1 High-throughput Technologies for Transcriptomic Profiling

#### *1.6.1.1 Microarrays*

Microarray technology was first explored for gene expression profiling in idiopathic PD in 2000 by Mandel and colleagues who investigated gene expression of dopaminergic neurodegeneration and neuroprotection in MPTP and 6-hydroxydopamine models (Mandel et al., 2000). Microarray technology continues to be utilized allowing for measurement of steady-state levels of mRNA transcripts derived from thousands of genes within a genome and for quantification of large amounts of messenger RNA (mRNA) transcripts simultaneously.

Briefly, utilization of microarray technology for transcriptomics involves the extraction of RNA and reverse transcription into complementary DNA (cDNA) which is then fluorescently labeled and hybridized to probes on the array (Schulze and Downward, 2001; Singh and Kumar, 2013). Commercially available microarrays can probe for 15000-30000 different human RNA molecules. The two most commonly utilized array platforms are cDNA and oligonucleotide arrays which are divided according to the material which is arrayed. cDNA arrays contain spots of probe DNA which consists of polymerase chain reaction (PCR) products amplified from DNA libraries or clone collections. Comparisons of two samples can be performed simultaneously using different dyes such as cyanine 3 (Cy3) and cyanine 5 (Cy5) making this technique useful for disease gene profiling. In contrast, oligonucleotide microarrays

use short 20-25 mers synthesized *in situ* and deposited by photolithography onto silicon wafers (these include high-density arrays from Affymetrix) or using ink-jet technologies.

Microarrays allow for high-throughput transcriptomic investigation at a relatively low cost. However, microarrays have a number of limitations. Firstly, they are based on known genome sequence data and therefore cannot detect novel transcripts. Hybridization reactions can be non-specific and they have small dynamic ranges due to background signals and signal saturation, making quantification of under or over expressed genes with changes of less than two-fold impractical (Wang et al., 2009).

Due to limitations involved in micoarray technology, RNA-Seq was developed as a high-throughput technology which can be used for transcriptomic profiling and could potentially overcome some of the pitfalls mentioned.

#### *1.6.1.2 RNA-Sequencing*

Since 2008, RNA-Seq has emerged as a leading transcriptomic technique as it allows for the detection of differential gene expression, determination of novel mRNA transcripts, alternative splicing events and the presence of long-non-coding RNAs (lncRNAs). In this approach, RNA is fragmented, then reverse transcribed into cDNA fragments where adaptors are ligated to one or both ends of the molecule (Figure 1.6). Each molecule is sequenced either from one end (single-end sequencing) or both ends (paired-end sequencing). Sequencing data are then aligned either to a reference genome, reference transcripts or *de novo* assembled to produce a transcription map or expression profile (Chu and Corey, 2012).

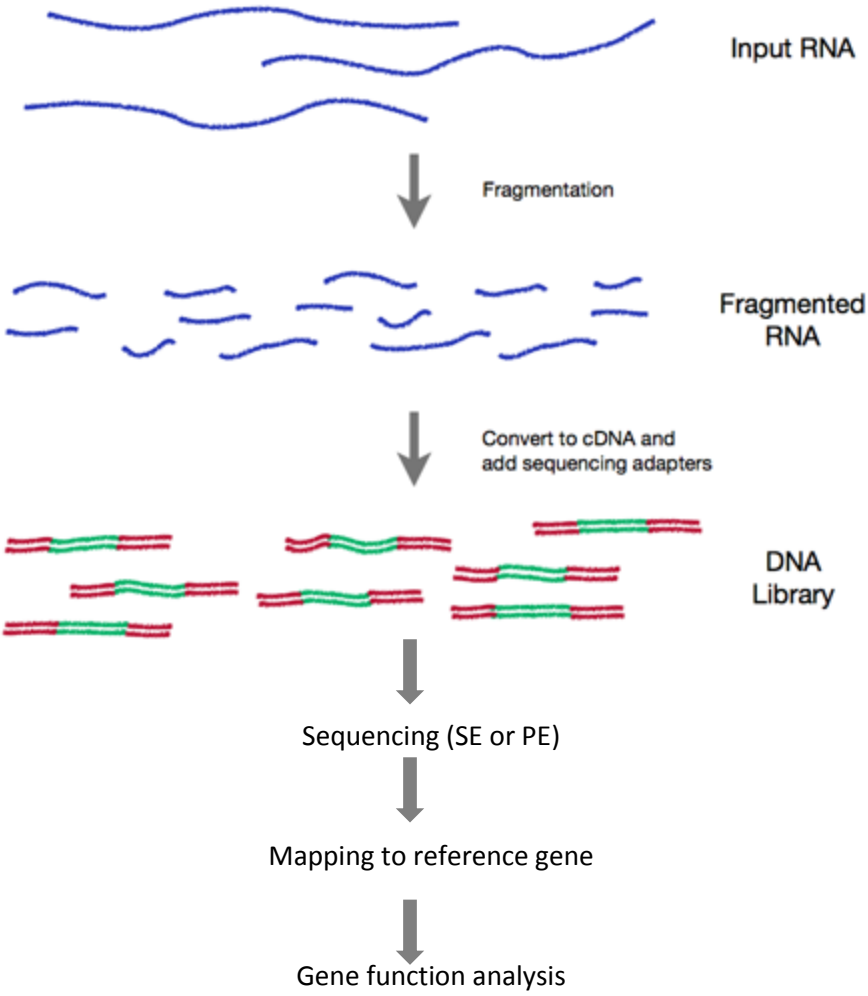



**Figure 1.6: RNA-Seq analysis workflow.** Total RNA is fragmented and reverse transcribed into DNA where adaptors are ligated to one or both ends of the molecule. The molecules are then sequenced either from one end (single-end ) or both ends (paired-end) and mapped to a reference genome. Gene function analysis is performed to produce an expression profile. SE, single-end; PE, paired-end. Adapted from http://rnaseq.uoregon.edu/

RNA-Seq offers many advantages as it allows for the investigation of novel transcriptomes. Information regarding transcript structure, including single base pair resolution of transcript boundaries and boundaries between exons can be revealed using this technique. Single nucleotide polymorphisms (SNPs) can also be detected allowing for genotyping and linkage analysis to be investigated. In RNA-Seq each molecule is sequenced individually and mapped to unique regions of the genome therefore overcoming problems with saturation and background signaling. RNA-seq is also highly reproducible and once data has been generated, it can be analyzed using a variety of pipelines to understand disease

events such as alternative splicing.

However, all new technologies have limitations and the challenges with RNA-Seq mainly lie in the bioinformatic analysis related to storage, retrieval, and the processing of the 'big data' generated. One of the initial challenges in data analysis is mapping of short reads to a reference genome, mainly with regards to the complexity of the human transcriptome but additionally selecting a reference that is suitable for the population being studied may prove more challenging than initially thought.

### 1.6.2 Summary of Previous Transcriptomic Studies in PD

The next part of this thesis will focus on published genome-wide transcriptomic studies that were performed in blood or brain tissue of PD patients and the main findings of each study will be highlighted. The first microarray-based gene expression analysis of PD in brain tissue was published in 2004 (Grünblatt et al., 2004) whereas the first study in blood was performed only in 2007 (Scherzer et al., 2007). Therefore, for the purposes of this review section, PD gene expression papers published prior to 2003 were not included. All relevant work published up to May 2016 were identified and reviewed (a total of 33 studies), and the genes and pathways highlighted have been summarized in Table 1.3. Studies excluded from Table 1.3 are listed with reasons in Appendix III, Table 3.

**Table 1.3: A summary of genome-wide mRNA and miRNA expression studies performed on blood and brain tissue from patients with Parkinson's disease**

| Reference | Tissue Type | Number of Samples | Platform and Kit Used | Number of Differentially expressed (DE) genes | Genes Highlighted | Pathways Highlighted |
|---|---|---|---|---|---|---|
| **Studies in blood** | | | | | | |
| Scherzer et al, 2007 | Whole blood | 50 PD patients<br>55 Controls | PAXgene procedure, RNA 6000 NanoChip kit on the Agilent 2100 Bioanalyzer. QIAquick purification kit, T7 MEGAscript reagents<br>**Microarray:**<br>Human Genome U133A array and scanned on an HP Gene Array scanner. | 22 unique genes<br><br>**Data results not in article:**<br>305 genes with FDR<0.2 and 228 genes by t test with p <0.005 | VDR, HIP2, CLTB, FPRL2, CA12, CEACAM4, ACRV1, UTX, *HSP70*, *ST13*, *SNCA* | Ubiquitin-proteasome pathway, mitochondrial function and apoptosis |
| Soreq et al, 2008 | Nucleated blood cells | 50 early stage PD patients<br>55 Controls | **Microarray:**<br>Affymetrix U133A 3' Array | 982 significant (p<0.05) probeset ID's identified differentiating between 3 states | *SNCA, DBH, ST13, ALAS2, SELENB, RPS4Y1, PTPRC, GNLY, ATRX, POLR2C, NK-TR* | Iron, regulation of apoptosis, mitochondrial depolarization, homeostatic processes, neuroinflammation. |
| Mutez et al, 2011 | PBMCs | 307 sporadic PD patients<br>158 familial PD patients<br>47 Controls | **Microarray:**<br>Agilent one-colour whole human genome 44K microarray, | 2280 differentially expressed probes in *LRRK2*-mutated subjects compared to controls (p<0.05). 827 annotated probes using GO annotation tools | *ALAS2, ARG1, DUSP10, ERAF, IFNG, LTF, SELENBP1, UTG2B17* | Axonal guidance signaling, TGF-β signaling, IL-10 signaling, leukocyte extravasation signaling, Parkinson's signaling, IL-4 signaling, neuregulin signaling, toll-like receptor signaling, IL-6 signaling, calcium signaling, death receptor signaling |
| Soreq et al, 2012 | Leukocytes | 7 PD patients<br>6 HC | **Microarray:**<br>Affymetrix | Highly alternatively spliced exons: 18 events | *AF130358.5, leucine-rich repeat containing* | Alternative splicing processes, neuroinflammation |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | GeneChip® Exon_1.0_ST_Array | in 18 distinct genes (2-fold SI change, SI p-value<0.005 and detection above background p-value<0.05. | *LRRC8C, LYPD6B, EP400NL, CLASP1, PDE3A, KIAA2026, TKTL1, ZNF257, DMTF1, MYCBPAP, USO1, AP1S2, VPS37A, CICP13, MYBBP1A, LA16c-60G3.8, DLGAP5* | |
| Alieva et al., 2014 | Whole blood | 4 PD patients<br>4 Controls | HumanHT-12 v4 Expression BeadChip Kit, | 1492 differentially expressed transcripts | *SNCA, DNM2, AP2, STX2, STX10, VAPA, VAMP4, VAMP8* | Immune system, cellular transport, α-synuclein, inflammation and alternative splicing |
| Mutez et al, 2014 | PBMCs | 9 PD patients (heterozygous LRRK2 G2019S mutation)<br>20 Sporadic PD patients<br>40 Controls | **Microarray:**<br>One colour whole-human genome 44K array (Agilent), Illumina Human HT-12 v4 Expression BeadChip | 367 genes deregulated in both platforms (Agilent and Illumina), | Four genes involved identified in all three studies: *ALAS2, EPB42, SELENBP1* and *SLC4A1* | AMPK signaling, inflammation, glucocorticoid receptor signaling, IL-4, IL-9, IL-15 signaling, T cell receptor signaling, CCR3 signaling, mitochondrial dysfunction, oxidative phosphorylation, ELF2 signaling, Rac signaling, deregulation of cancer signaling, cell survival, and P13K/AKT signaling and purine metabolism. |
| Chikina et al, 2015 | Whole blood | 34 PD patients<br>32 healthy controls<br><br>20 male mice samples | 113 marker panel | Five novel blood markers (DHX58, IL1B, KDM68, PYCARD, TGFB1) significant changes in both the mouse and human species | *MyD88, DHX58, PYCARD IL1B, TGFB1, KDM6B, CASP9, MAPK3, DICER1, MAPK1, ARG1, PRKCH, GPR18, CD88, CD8A, TARBP2* | Innate and adaptive immunity, inflammation |
| Santiago and Potashkin, 2015 | Whole blood | **Microarray datasets:**<br>83 PD patients<br>46 Controls<br><br>**Clinical samples:** | Dependent on the individual datasets. | 2,781 differentially expressed genes across all four datasets (680 up-regulated, 2101 down-regulated) | *THY1, HNF4A, PTBP1, UBC* | Bacterial invasion of epithelial cells, mitogen-activated protein kinase-signaling pathway, fructose and mannose metabolism, T-cell receptor signaling pathway, mammalian target of rapamycin-signaling |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 101 PD patients<br>91 Controls | | | | pathway, type2 diabetes mellitus and colorectal cancer. Down-regulated genes in blood of PD associated with KEGG pathways; protein processing in ER, Epstein-Barr virus infection, cancer types including prostrate, endometrial and lung cancer, UBC and PTBP1 |
| Infante et al, 2016 | Whole blood | 20 PD patients (G2019S *LRKK2* mutation)<br>20 asymptomatic carriers of the G2019S mutation<br>20 idiopathic PD patients<br>20 controls | Illumina HiSeq sequencing system (20 million paired end reads) | 2982 DEGs after adjusting for multiple testing | *HBD, HBG1, HBM, DGKQ, TMEM175, MMRN1, SREBF1, PRSS53, SYT11* | **<u>Common pathways in idiopathic PD and LRRK2 G2019S:</u>**<br>Complement and coagulation cascades (KEGG and reactome), cell adhesion molecules (CAMs), hematopoietic cell lineage (KEGG), and extracellular matrix (ECM)<br>organization (Reactome).<br><br>Pathways common to LRRK2 G2019S:<br>Interpheron alfa/beta signaling, O2/CO2 exchange in erythrocytes and NCAM1 interactions |
| **Studies in brain tissue** | | | | | | |
| Grünblatt et al, 2004 | SN and cerebellum | 6 PD patients<br>6 Controls | **Microarray:** Affymetrix HG-FOCUS oligonucleotide array. | 137 DEGs, Stringent analysis (1.5 FC in five out of six patients) resulted in 20 DEGs (8 down-regulated, 12 up-regulated) | *ARPP-21, VMAT2, ALDH1A1, SKP1A, HSPA8, SRPK2, TRIM36, TMEFF1, PARVA, LGALS9, SELPLG, PENK, LRP6, EGLN1, MAN2B1, SPHK1, LOC56920, ZSIG11, SRRM2* | **Down-regulated genes:** signal transduction, protein degradation, dopaminergic transmission/metabolism, ion transport, protein modification, and phosphorylation and energy pathways.<br>**Up-regulated genes:** cell adhesion, extracellular matrix |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | components, cell cycle, protein modification, phosphorylation, protein metabolism, transcription and inflammation and stress. |
| Hauser et al, 2005 | SN | 6 PD patients<br>2 PSP<br>1 FTDP<br>5 Controls | **Microarray:**<br>Affymetrix® human genome U133A microarrays | 142 DEGs between PD compared to controls | *HSPA1A, HSPA1B, UCHL1, parkin, COX41, ATP1B1, STXBP1* | **Chaperones, ubiquitination, vesicle trafficking and nuclear-encoded mitochondrial genes** |
| Miller et al, 2005 | SN and STR | 6 PD patients<br>8 Controls | CodeLink Human UniSet 20K bioarray | 19800 DEGs, 323 DEGs in PD SN compared to control | *DAT, EN-1, AADC, SYNGR3, NSF, STX1A, STXBP1, SYT1, MARK-1, MAP-2, DSCR1L1, VEGF, GSTM2, UCHL-1, HK1, MMP23B, SREBP-2* | Synaptic pathways, cytoskeletal maintenance, UPS, anti-oxidant and detoxification, nigrostriatal pathway |
| Zhang et al, 2005 | BA9, PUT and SN | 15 PD patients<br>15 Controls | **Microarray:**<br>Human Genome U133A GeneChip® (Affymetrix®) | 329 DEGs | *RTP801, HSPB1, MKNK2, FGF13, SNAP25, RGS4, SMA5* | Ubiquitin proteosomal system, electron transport chain, MAPK pathway, heat-shock proteins, lysosomal pathway, apoptotic mechanisms |
| Papapetropoulos, 2006 | 21 brain regions: SN, VTA, BA35, IC, AMYG, NBM, CN, PUT, NAc, GPI, MD, PI, STH, DMNV, CH, ACV, DRN, LC, HTH, HPC, RPRF | 22 PD patients<br>23 Controls | **Microarray:**<br>Human Genome U133 Plus 2.0 GeneChip array | 11 candidate genes | *MRPS6, HIST1H2BD, RBM3, SLC38A2, CHORDC1, CIRBP, FLJ33814, FUSIP1, PRKACB, STIP1, SUV420H1* | Mitochondrial dysfunction and cAMP/protein kinase A signaling, apoptosis, cell signaling, and cell cycle |
| Vogt et al, 2006 | PUT, OC, cerebellar hemispheres | 8 PD patients<br>8 MSA patients<br>8 Controls | **Microarray:**<br>Affymetrix® human U133A GeneChip®Arrays | **<u>Putamen:</u>**<br>Bonferroni correction: 40 DEGs (20 down and 20 up-regulated)<br><br>Benjamini-Yekutieli correction: 84 DEGs in PD | Neuropeptide Y, heterogenous nuclear ribonucleoprotein C1/2, *HTR2C, SCA1*, glutathione S-transferase theta 1 and 2, cytochrome P450, | Transcriptional regulation, G-protein-coupled receptor signaling pathways, RNA binding, protein modification and folding, intracellular signaling cascades |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | compared to controls<br><br>Four genes regulated in both MSA and PD<br><br>**<u>Cerebellum:</u>**<br>88 genes differentially regulated (49 down and 39 up-regulated)<br><br>12 genes regulated in both MSA and PD | family 1, subfamily B, polypeptide1 and somatomedian A, *GALNT12, PPAT, DACT1, STK10, ADORA2A, DRPLA, NR4A2, KIAA00472, CDR1.* | |
| Duke et al, 2007 | SN (lateral and medial) | 9 PD patients<br>7 Controls | **Microarray:** Affymetrix HU_133A and HU_133B gene chips | 396 sequences DE between SNl and SNm ($p<0.01$), more stringent criteria: 24 DEGs and 10 expressed sequence tags | *CD30*, TNF receptor associated factor 4, *TNF-$\alpha$* induced protein 3, IL-5 receptor $\alpha$, IL-6, IL-8 receptor $\beta$, NDUF$\alpha$6, NDUF$\alpha$2,NDUF$\alpha$9, programmed cell death 1, BCL2-like 11, caspase 2, caspase 8, caspase 10. | TNF and IL signaling pathways, pro-inflammatory cytokines, mitochondrial complex 1, inflammation, oxidative stress. |
| Grünblatt et al., 2007 | HPC, CE and Gfm | 9 PD<br>14 AD<br>9 Controls | **Microarray:** Affymetrix HU_133A and HU_133B gene chips | 3 DEGs in all three brain regions and all three diseases: *SNX2, IRS4* and *H1ST1H3E* | *SNX2, IRS4, H1ST1H3E, SYT1, OR10H3, SNX2, h5HT4b, APBA2, VPS35, VPS41, GSTM1, COG2, IRS4* | Signal transduction, energy metabolism, stress response, synaptic vesicle synthesis, calcium binding, oxidative stress |
| Moran and Graeber 2008 | SN (medial and lateral) | 15 PD patients<br>8 Controls | **Microarray:** Affymetrix HU_133A and HU_133B gene chips | 892 highly dysregulated priority genes. Up-regulation of SN genes | *ABCA8, ARGHEF10, ARRDC2, BBX, BOC, C10orf104, C10orf128, C10orf54, C22orf9, CHORDC1, CLEC2B, CRLF3, DDA3, DEPP, DKFZp667G2110, DNAJB6, DOCK5, ELTD1, FAM123A, FLJ11280, FLJ40092,* | Inflammation, PINK1, Parkin, UPS pathway. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | *GUSBP1, H2AFL, HIGD1B, IFITM2, KCNMB4, KIAA0500, KST1, LOC399959, LOC402573, LOC440248, MGC22265, MKNK2, mn7, MRC1L1, MT1H, MT1K, NSUN6, PFAAP5, PPAP2C, PRKY, RPS11, SEC14L1, SLC04A1, TncRNA, TRIM4, TT1D, USP34, USP52, USP54, ZDHHC11, ZNF302* | |
| Elstner et al, 2009 | FC and mid-brain sections | 8 PD patients<br>9 Controls | Illumina WG6v1 expression chips | 8491 transcripts, 4 genes validated | *PDXK, TRAPPC4, ND2, SRGAP3, MTND2* | Mitochondrial respiratory chain, protein trafficking, mitochondrial dysfunction, vitamin B6, dopamine metabolism, axon guidance, vesicle transport. |
| Simunovic et al, 2009 | SNc, isolation of DA neurons | 10 Idiopathic PD cases<br>9 Controls | **Microarray:** Affymetrix HU-133A arrays | 580 up-regulated and 465 down-regulated genes | *RAP1GAP, UCHL1, RIMS3, ATP13A2, Parkin, PINK1, RIMS1, LRRK2, DJ-1, SLC6A3, UBE2K, TH, KCNJ6, SNCA, GUSB* | Genes associated with PCD, mitochondrial function, protein degradation, synaptic function, growth factors, receptors and ion-channels.<br>PARK gene, programmed cell death |
| Bossers et al., 2009 | Brain tissue | 7 PD patients<br>8 Controls | RNAlater®-ICE, Qiagen RNeasy Mini Kit™, NanoDrop® ND-100 spectrophotometer, Agilent™ 2100 bioanalyser, Agilent custom-made 22K 60-mer oligonucleotide arrays, Agilent™ DNA Microarray Scanner, | Differential expression of 287 transcripts in the SN, 16 transcripts in the caudate nucleus and 4 transcripts in the putamen | **SN:**<br>*SYT1, TRIM36, MDH1, NSF, SASH1, ALDH1A1, DSCR1L1, SNX10, SNCA, NURR1, AGTR1, VMAT2, PTS*<br>**Caudate nucleus:**<br>*LTF, CYBRD1*<br>**Putamen:**<br>*GNPDA2, SIPA1L1, ADM* | Neurotrophin signaling, GDNF signaling pathway, axon guidance, iron-binding protein lactotransferrin (LTF) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | ABI 7300 sequence detection system | | | |
| Stamper et al, 2009 | Posterior cingulate cortex | 15 PD<br>14 PDD<br>14 Controls | **Microarray:**<br>Human Genome U133 plus 2.0 arrays | 162 significantly altered genes between PD patients and controls.<br>556 genes altered in PDD cases compared to controls and PD patients.<br>69 genes in PD cortical neurons which occur before the onset of dementia | *PSMB4, SLIT2, FGF9, TIM50L, SSBP, SART3, LUC7L, FNBP3, PLRG1, FUS* | Oxidative stress, axonal transport, neurite outgrowth, axonal pathfinding, synaptic transmission<br>Pre-mRNA splicing |
| Simunovic et al, 2010 | SNc | 10 idiopathic PD<br>9 Controls | **Microarray:**<br>Affymetrix HU-133A array | 642 down-regulated, 53 up-regulated gen | Gender specific expression profiles: *ATP5G3, ATP6V1E1,CHRNA4, CLTC,COX7C,DRD2, NDUFB2, PRKACB, SLC35A1, SLC6A1, TIMM44, UQCRH, ZNF606,* oxidative phosphorylation, apoptosis, synaptic transmission, transmission of nerve impulse, *HSP90AA1, HSPA8* | Oxidative phosphorylation, apoptosis, synaptic transmission, UPS, mitochondrial pathways, neurotransmitter and ion channels |
| Botta-Orfila et al, 2012 | Locus coeruleus (LC) | 9 PD patients<br>5 Controls | GeneChip Human Exon 1.0 ST Array | 232 differentially regulated genes | *C3, HLA-DPA1, HLA-DQA1, HLA-DRA, NEUROD1, SLC5A7* | Synaptic transmission, cell signaling cascades, structural and synaptic plasticity processes, LTP and GABA receptor signaling and calcium signaling pathways |
| Corradini et al, 2014 | Vagal nucleus (VA), locus coeruleus (LC), SN | 8 Idiopathic PD patients<br>7 Controls | 44K DNA microarrays | DEGs:<br>VA: 234 genes<br>LC: 183 genes<br>SN: 326 genes. | *VA (CT:PD): FLYWCH1, HRC, S100A4, PGM3, SOX10, CPNE2, LOC100128821, AGBL4, IFT88, ADAM15, SHARPIN,* | VA-CT hubs: Ca2+ homeostasis, myelination, neuroprotection in aging<br>VA-PD hub:<br>Proteotoxic stress, PARK14 (PLA2G6). |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | *GNL3L, UBTF, ARS2, PLA2G6*<br><br>*LC (CT:PD): GPRC5B, GRM3, UGT8, NUDT13, SEPP1, RGS5, PCOLCE2, PPP4R1, FAM5B, MED30, TOB2, ZNRF3, PARP4, ATXN1, SFRS18, GAS7*<br><br>*SN (CT:PD): SIRT1, ZEP112, SHC4, TMEM123, BCKDHB, CBFB, CLDND1, GLDN, MBTD1, ARID4B, HNRNPA3, SORT1* | LC-CT hubs: Protection against oxidative and proteotoxic stress, myelination and BBB maintenance<br>LC-PD hubs: Neuroprotection and brain homeostasis, SFRS18, ZFNR3, FAM5B<br>SN-CT hubs: Neuroprotection and homeostasis in the aging brain, SIRT1, BCKDHB<br>SN-PD hubs: Neurodegenrative processes |
| **Studies in blood and brain tissue** | | | | | | |
| Dusonchet et al, 2014 | Whole blood, Substantia nigra pars compacta (SNpc) and the frontal cortex | **Whole blood:**<br>50 PD and 22 control cases from GSE8397<br><br>**Brain tissue:**<br>29 PD and 18 control tissues from GSE6613: 15 medial parkinsonian SN, 9 lateral parkinsonian SN, 8 medial nigral control, 7 lateral nigra control, 5 frontal cerebral cortex PD, 3 frontal cerebral | **Microarray:**<br>Affymetrix Human Genome U133A and U133B gene chips | 200 genes from CLR network | *LRRK2, VPS34, PIK3C3, HDAC6, PARKIN, PINK1, MAP2K7, MAP2K2, MAPK1, MAP2K6, MAP2K7, MAP2K8, CDK5, FZD1, WNT4, FZD9, RGS2* | Protein translation, cytoskeletal processes, vesicular dynamics, mitochondrial damage and autophagy |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | cortex control | | | | |
| Riley et al, 2014 | Whole blood, Cortex, Stratium and the Substantia nigra | **Brain region samples:**<br>**Substantia nigra:**<br>16 PD samples<br>14 normal<br>**Stratium:**<br>16 PD samples<br>16 normal<br>**Cortex:**<br>15 PD<br>19 normal<br>**Non-PD neurodegeneration**<br>6 cortex<br>6 stratium<br><br>**Blood Samples:**<br>46 whole blood lysate samples, 24 Controls, 6 non-PD neurodegeneration, 16 PD samples | **Microarray:**<br>Affymetrix GeneChip Sense Target (ST) arrays, Affymetrix HG U133 Plus 2.0 platform<br>**RNA-Seq:**<br>QuantiGene Plex 2.0 Human 57-plex panel | None | *CXCR4, CAMK2, Aldha1, USP9Y, HBB, RELN, ADORA2A, FOS, SLC04A1, DDT1T4, PENK, ALDH1A1, TH, FABP7, PCDH8, MAG, OPALIN, DDX3Y, VCAM1, EDN1, P2RX7, SLC17A8, NLGN4Y, GRIA1* | Overlap in pathways using microarray and RNA-Seq<br>**Up-regulation:**<br>**Stratium:** Neuron survival and death in AD, D1A signaling receptor, CREB pathway, Neuron survival and neurogenesis in bipolar , NMDA-dependant postsynaptic long-term potentiation in CA1 hippocampal neurons<br>**SN:**<br>HSP70 and HSP40-dependant folding in HD, WNT signaling, CFTR folding and maturation, protein aggregation in HD, vitamin K metabolism<br>**Cortex:**<br>Neuron survival and neurogenesis in Bipolar disorder, classical complement system, cytoskeleton in oligodendrocytes differentiation and myelination, ERK1/2 signaling pathway, neurodevelopment in Schizophrenia.<br>**Down-regulation:**<br>**Stratium:**<br>Interactions in oligodendrocyte differentiation and myelination, inhibition of remyelination in multiple sclerosis, MAG-dependent inhibition of neurite outgrowth, role of thyroid in regulation of oligodendrocyte differentitation, role of thyroid in regulation of oligodendrocyte |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | differentitation in multiple sclerosis<br>**SN:**<br>G-protein-coupled receptors signaling in lung cancer, WNT signaling in HCC, Slit-Robo signaling, RAB3 regulation pathway, cytoskeleton in oligodendrocyte differentiation and myelination<br>**Cortex:**<br>Regulation of EMT, TGF-beta 1-mediated induction of EMT in epithelium, role of AP-1 in regulation of cellular metabolism, IL-1 signaling in melanoma, MIF-mediated glucocorticoid regulation. |
| **Studies investigating miRNAs** | | | | | | |
| Margis et al, 2011 | Peripheral blood | 8 NT PD<br>7 EOPD<br>8 Controls | Applied Biosystems StepOne System | Three miRNAs reduced profiles of relative expression | miR-1, miR-22*, miR-29a, miR-16-2*, miR-26a2*, miR30a, alpha-synuclein, TPPP/p25, FGF20, TP53BP2, CDC42, cell cycle, BDNF, IGF-1. | None highlighted |
| Martins et al, 2011 | PBMCs | 19 PD<br>13 Controls | Exiqon-developed miRCURY™ LNA microarray | 18 differentially expressed miRNAs , 662 predicted target genes | miR-301a, miR-19b, miR-126*, miR-335, miR-29b, miR-29c, miR-30b, miR-374a, miR-126, miR-199a-5p, miR-199a-3p/miR-', miR-28-5p, miR-151-5p, miR-151-3p, miR-147, miR-30c, miR-26a, miR-374b, $\alpha$-synuclein, USP37, *ST8SIA4,* Inflammation, Immune | ubiquitin proteasome system, chaperone-mediated autophagy, glycosphingolipid biosynthesis, protein ubiquitination path-way, emaphorin signaling, DNA methylation and transcriptional repression signaling, retinoic acid receptor (RAR) activation, hypoxia signaling in the |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | system, ubiquitination | cardiovascular system, O-glycan biosynthesis, protein ubiquination pathway, synaptic long-term potentiation and nicotinate and nicotinamide metabolism |
| Botta-Orfila et al, 2014 | Serum | 95 IPD<br>21 LRRK2 PD<br>95 Controls | **Discovery study:** TaqMan Low Density Array (TLDA) on the Viia7 1.0 Real-time PCR<br>**Validation study:** TaqMan MicroRNA assays using Viia7 1.0 Real-time PCR system<br>**Second Validation Study:** individual TaqMan MicroRNA assays using StepOnePlus Real-time PCR system | Three miRNAs with decreased levels in serum in IPD and LRRK2 PD patients compared to controls. | miR-29c, miR-29a and miR-19b | ECM-receptor interaction, focal adhesion, MAPK, Wnt, mTOR, adipocytokine and neuron projection pathways |
| Chatterjee et al, 2014 | PBMCs | 19 PD patients<br>13 Controls | Microarray expression data from GEO | 9 novel IR hub miRs | hsa-miR-200c, has-miR-200b, has-miR-200a, has-miR-17, has-miR-19a, has-miR-20a, has-miR-18a, has-miR-141, has-miR-92a. | KEGG pathways (axon guidance, ubiquitin mediated proteolysis, oocyte meiosis, focal adhesion, TGF-β signaling, MAPK signaling pathway, small cell lung cancer, dilated cardiomyopathy and GnRH signaling pathway), cancer pathways, cardiovascular disease pathways |
| Serafin et al., 2014 | Whole peripheral blood | 38 idiopathic PD patients<br>38 controls | TaqMan miRNA Reverse Transcription kit, Bio- | Two reference genes identified | RNU24, Z30, miR-103a-3p, miR-29a-3p, miR-30b-5p | None |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | Rad 96CFX instrument | | | |
| Soreq et al, 2014 | Blood leukocytes and the Amygdala and SN | **Blood Leukocytes:** 3 PD patients 3 Healthy controls **Brain Samples:** 5 PD patients 5 Healthy controls | **RNA-Sequencing:** Applied Biosystems SOLiD-3 platform | RT-PCR validations in PD leukocytes and the amygdala and SN | U1 spliceosomal lncRNA, RP11-462G22.1 | Oxidation-reduction signaling pathways, mitophagy, protein metabolism, oxidative stress, mitochondrial functioning, JNK signaling pathway, immune system pathways, |

PBMCs, peripheral blood mononuclear cells; DEGs, differentially expressed genes; FDR, false discovery rate

#### *1.6.2.1 Gene Expression Studies Performed in Blood*

There have been 17 transcriptomic studies on blood in the field of PD, and these are summarized in Table 1.3. Nine of these studies were on whole blood, four on peripheral blood mononuclear cells (PBMCs), one on nucleated blood cells, one on serum and two on leukocytes.

##### Samples Used

Sample numbers vary greatly across the studies from a minimum of 3 PD patients (Soreq et al., 2014) to larger cohorts consisting of 465 PD patients (Mutez et al., 2011). The average number of PD samples analyzed was 62.6, with a median of 20. The majority of studies used the United Kingdom Parkinson's Disease Society Brain Bank criteria for PD diagnosis. Additionally, depending on the research question posed, other criteria for disease diagnosis included assessment of disease severity using the Unified Parkinson's Disease Rating Scale in two studies (Margis et al., 2011; Martins et al., 2011) and Hoehn and Yahr staging in two studies (Martins et al., 2011; Scherzer et al., 2007). Five studies utilized known mutation carriers of the *LRRK2 G2019S* mutation (Chikina et al., 2015; Infante et al., 2015; Mutez et al., 2014, 2011; Santiago and Potashkin, 2015).

The basis for most gene expression studies is case–control comparison and therefore selection of control samples is of great importance. Variables such as age, sex and ethnicity between controls and patient samples were well controlled for in the majority of the reviewed studies. Control sample numbers in these studies ranged from 3 (Soreq et al., 2014) controls to 95 PD controls (Botta-Orfila et al., 2014). The average number of controls analyzed in blood studies was 33.6, with a median value of 22. Only two studies out of the 17 specifically stated that the controls were healthy (Mutez et al., 2011; Soreq et al., 2008).

Controls were also matched for sex in 7 studies, ethnicity in 2 studies and age in 11 of the 17 studies. Another criterion for suitable control selection in some studies included no personal or family history of neurological disease (Chikina et al., 2015; Scherzer et al., 2007a).

##### Platforms Used

Some of the oligonucleotide arrays utilized in blood gene expression studies on PD included the Affymetrix Human Genome U133A array (Scherzer et al., 2007a; Soreq et al., 2008) and the miRCURY™ LNA microarray (Martins et al., 2011).. The Agilent Human Genome 44K microarray (Mutez et al., 2014, 2011) was the only cDNA microarray platform utilized in the summarized studies. Some studies performed meta-analysis using DNA microarray datasets that were obtained from the NCBI Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/) and the Gemma database

(http://www.chibi.ubc.ca/Gemma/home.html) which differed in the platform utilized either being Affymetrix Human Genome U133A, U133B or both gene chips depending on the dataset. The GEO dataset with accession number GSE6613 was investigated in two meta-analyses using the Affymetrix Human Genome U133A gene chip (Dusonchet et al., 2014; Santiago and Potashkin, 2015). Other datasets used were GSE8397 (Dusonchet et al., 2014), GSE22491, GSE22491, GSE54536 (Santiago and Potashkin, 2015) and GSE16658 (Chatterjee et al., 2014a).

In one study, exon arrays were used to identify alternative splicing events in PD patients compared to healthy controls (Soreq et al., 2012). Only two studies used RNA-Seq technique, Riley and colleagues used the Human 57-plex panel (Riley et al., 2014) whereas the second study used the SOLiD-3 platform (Soreq et al., 2014).

#### Differentially Expressed Genes and Pathways Identified using PD Blood

Differential gene expression is the main focus of transcriptomics and in total the number of differentially expressed genes identified ranged from 5 novel blood biomarkers (Chikina et al., 2015) to 2,982 (Infante et al., 2016).

In the five studies that investigated differential gene expression in PD patients with the *LRRK2 G2019S* mutation, there was little concordance between the genes and pathways identified (Chikina et al., 2015; Infante et al., 2016; Mutez et al., 2014, 2011; Santiago and Potashkin, 2015). One possible explanation could be the fact that the studies used different platforms for investigation.

Differential gene expression findings from meta-analyses are highly dependent on the accuracy and completeness of each individual dataset and therefore heterogeneity can skew data. Varying statistical analysis utilized in each meta-analysis can also result in differences in the genes identified. Interestingly, no overlapping genes or pathways were identified when comparing two different meta-analyses which utilized the same microarray dataset (GEO accession number: GSE6613) (Dusonchet et al., 2014; Santiago and Potashkin, 2015).

### *1.6.2.2 Gene Expression Studies Performed on Brain Tissue*

PD is a neurodegenerative disease and therefore studying the affected region of the brain directly is necessary. Several genome-wide expression studies on brain tissue of PD patients have been carried out since the first published study in 2004 (Grünblatt et al., 2004). A total of 19 transcriptomics studies on brain tissue are summarized in Table 1.3. The majority of these studies primarily investigated the SN, which is the primary region associated with PD in which DA neuronal death occurs (Bossers et al., 2009;

Cantuti-Castelvetri et al., 2007; Corradini et al., 2014; Duke et al., 2007; Hauser et al., 2003; Lu et al., 2006, 2005, Simunovic et al., 2010, 2009).

### Samples Used

Sample sizes ranged from 5 (Soreq et al., 2014) PD brain samples to 29 (Dusonchet et al., 2014). A mean of 13.6 PD samples were used with a median of 9 samples.

The mean number of control samples was 11.7 (medium of 9), with a range from 5 (Hauser et al., 2005; Soreq et al., 2014) to 23 ( Papapetropoulos, 2006). Control brain samples were obtained from healthy individuals in 2 studies (Botta-Orfila et al., 2014; Grünblatt et al., 2004), age-matched in 6 studies (Bossers et al., 2009b; Duke et al., 2007; Grünblatt et al., 2004; Simunovic et al., 2010, 2009; Stamper et al., 2008) and sex-matched in 5 studies (Bossers et al., 2009b; Simunovic et al., 2010; Zhang et al., 2005). Some studies also matched post-mortem delay (PMI) (Bossers et al., 2009b; Duke et al., 2007; Grünblatt et al., 2004; Simunovic et al., 2010, 2009; Zhang et al., 2005) , and brain pH (Bossers et al., 2009b; Duke et al., 2007; Zhang et al., 2005) of the controls to PD patient samples. Interestingly, only one study matched brain weight of controls to PD patients (Bossers et al., 2009b).

Several studies have investigated other brain regions which are in close proximity to the SN and could potentially be involved in PD pathophysiology. These regions include the cerebellum (CE) (Grünblatt et al., 2004), putamen (PUT) (Zhang et al., 2005), Broadmann's Area 9 (BA9) (Zhang et al., 2005), cerebral cortex (Riley et al., 2014), frontal cortex (Dusonchet et al., 2014) and the stratium (Riley et al., 2014).

### Platforms Used

Oligonucleotide microarrays were used most extensively for genome-wide expression analysis in brain tissue with the Affymetrix Human Genome U133A (Duke et al., 2007; Dusonchet et al., 2014; Moran and Graeber, 2008; Simunovic et al., 2010, 2009; Vogt et al., 2006; Zhang et al., 2005) and U113B (Duke et al., 2007; Dusonchet et al., 2014; Moran and Graeber, 2008) being the main platforms. Other oligonucleotide arrays include the Agilent custom made 22K 60-mer array (Bossers et al., 2009b) and the Affymetrix Human Genome U133 Plus 2.0 array(Riley et al., 2014; Stamper et al., 2008). Only one study used the Human genome 44K microarray for expression analysis (Mutez et al., 2014) a cDNA microarray platform. Notably, RNA-Seq has been utilized as a technique in only two studies; on the Human 57-plex panel (Riley et al., 2014) and SOLiD-3 platform (Soreq et al., 2014) in both blood and brain.

### Differentially Expressed Genes and Pathways Identified using PD Brain Tissue

Studies utilizing brain tissue from multiple regions revealed region-specific differentially expressed genes (Bossers et al., 2009b; Riley et al., 2014; Vogt et al., 2006).

The differentially expressed genes belonged to multiple biological pathways including signal transduction (Grünblatt et al., 2004), protein degradation (Grünblatt et al., 2004; Simunovic et al., 2009), ion transport (Grünblatt et al., 2004), ion metabolism (Mutez et al., 2014), ubiquitination (Noureddine et al., 2005; Zhang et al., 2005), electron transport chain, apoptosis, mitogen-activated protein kinases (MAPK) pathway (Zhang et al., 2005), pro-inflammatory cytokines (Duke et al., 2007) and tumour necrosis factor (TNF) signaling (Duke et al., 2007).

Three of the studies reviewed found differential expression of heat shock proteins, which are molecular chaperones that mediate protein degradation. Grünblatt and colleagues identified decreased expression of HPS8 which encodes for a 70 kDa heat-shock cognate protein (Hsc70), a member of the heat-shock protein 70 (Hsp70) family (Grünblatt et al., 2004). Subsequently, a genetic association study demonstrated increased PD susceptibility with a polymorphism in the 5' promoter region of *HSP70-1*. Down-regulation of heat shock and associated proteins was observed, specifically suppression of tumorigenicity 13 (*ST13)*, a cofactor of HSP70 (Simunovic et al., 2009). Furthermore, Simunovic and colleagues identified six genes associated with protein degradation including heat shock protein 90 kDa alpha (HSP90AA1) and heat shock 70 kDa protein 8 (HSPA8) which were downregulated in males (Simunovic et al., 2010).

Genes associated with mitochondrial dysfunction have been shown to have increased expression in five studies (Duke et al., 2007; Dusonchet et al., 2014; Mutez et al., 2014; Noureddine et al., 2005; Simunovic et al., 2009).

#### *1.6.2.3 Gene Expression Studies of miRNAs*

Decreased levels of three miRNAs were found in the serum of idiopathic PD and PD patients with leucine-rich repeat kinase 2 (*LRRK2)* mutations compared to controls in one study (Botta-Orfila et al., 2014). Interestingly, miR-29a showed decreased expression in *de novo* PD patients (Margis et al., 2011) and serum of idiopathic PD and *LRKK2* PD patients (Botta-Orfila et al., 2014). Increased expression of miR-29a has been shown to indirectly be involve in the interaction between natural killer cells and interferon-γ (IFN-γ) and the dopamine receptor (Mikulak et al., 2014). MiR-29c and miR-19b were found to be differentially expressed in two different studies (Botta-Orfila et al., 2014; Martins et al., 2011) using two different platforms; Exiqon-developed miRCURY™ LNA microarray and TaqMan Low Density

Array respectively which provides evidence that investigation of miRNAs may be of relevance for PD pathobiology.

#### 1.6.2.4 Our Analysis of the Published Literature

To further investigate differentially expressed genes in these studies, we systematically and comprehensively compiled the gene list from each study. A total of 4,199 differentially expressed genes were identified across all studies using blood as a tissue source (Appendix I, Table 1). Analysis of studies using brain tissue resulted in a total number of 12,117 differentially expressed genes being identified (Appendix II, Table 2). Studies using the same dataset were excluded from the analysis as this could result in skewing of the results. All studies that were excluded are listed in Appendix III, Table 3. Micro RNA (miRNA) studies were also included in our analysis (Appendix IV, Table 4).

Next, we compared the gene lists to identify overlapping genes. Genes identified in two or more studies were considered as 'overlapping'. The last column of the blood and brain tables labelled as 'the grand total' provides the number of times that the gene is found to be differentially expressed across the various studies. Therefore, if the value in the grand total column is 2 this would indicate that that specific gene was found to be differentially expressed in two different studies. These studies can then be identified by the top row which includes the first author and date of publication. A summary table of all studies included in the overlapping gene lists is provided in Appendix V, Table 5. Therefore, the information and overlapping gene lists provided in Appendix I and II, Table 1 and 2 are an important resource for future gene expression studies in PD and were used for prioritization of candidate genes in the present study.

## 1.7 The Present Study

### 1.7.1 Research Problem Statement

The study of gene expression profiles in PD patients may lead to a better understanding of the pathogenic processes underlying this disorder. Studying and comparing the entire transcriptome between PD patients and controls may reveal novel insights into why brain cells die in these patients.

The focus of the present study was to study gene transcription at the level of the entire transcriptome using the RNA-seq technique to identify dysregulated pathways in a group of South African PD patients. We selected the RNA-seq technique as it has many advantages over other techniques including having the ability to measure both expression levels and RNA modifications (e.g. alternative splicing), detection

of low abundant transcripts, and the possibility of discovery of novel transcripts (Lewis and Cookson, 2012).

Patients used for the present study form part of a larger parent study entitled: Understanding the Shared Roots of Neuropsychiatric Disorders and Modifiable Risk Factors for Cardiovascular Disease. This parent study, hereafter referred to as 'Shared Roots', involves the recruitment of 600 individuals with a neuropsychiatric disorder (200 with PD, 200 with schizophrenia, 200 with post-traumatic stress disorder) and 600 matched controls. The main aim of Shared Roots is to interrogate genomic, neural, cellular and environmental signatures that are common to neuropsychiatric disorders and cardiovascular disease risk to unravel mechanistic pathways that contribute to co-morbidity, symptom severity, and treatment outcomes.

Study participants were exposed to extensive clinical assessments and biological sampling including bloods (for genomic DNA and RNA isolation), skin biopsies (for culturing of fibroblasts) and hair sampling.

All of the study participants are from the mixed ancestry ethnic group and reside within Cape Town in South Africa. This ethnic group is only found in South Africa and comprises a unique admixture of Khoesan (32-43%), Bantu-speaking Africans (20-36%), European (21-28%) and Asian (9-11%) populations (de Wit et al., 2010). The unique genetic make-up of this population group makes them ideal for genetic studies, and also the environmental influences (climate, socioeconomic circumstances etc.) that are distinct to South Africa makes it important to study gene expression in these patients to determine whether they are the same or different to the findings indentified in European, North American Caucasian and Asian populations.

We first undertook to identify whether any of the patients harbored PD-causing mutations as the disease process in these patients may be different to the pathways implicated in idiopathic PD (Mutez et al, 2011). To this end, patients were screened for the most common PD-causing mutation, G2019S in the *LRRK2* gene, as well as the A30P mutation in the *SNCA* gene and exonic rearrangements in all of the known PD-causing genes. This was achieved using commercially-available kits based on the multiplex ligation-dependent probe amplification (MLPA) technique. Thereafter, whole genome expression profiling was applied to this group of patients and controls.

### 1.7.2 Hypotheses

**Null Hypothesis (Ho):** Transcriptomic profiling using RNA-Seq in a subset of study participants from the Shared Roots study will not reveal significant differences in gene expression and pathways between PD patients and controls

**Alternative Hypothesis (Ha):** Transcriptomic profiling using RNA-Seq in a subset of study participants from the Shared Roots study will reveal significant differences in gene expression and pathways between PD patients and controls

### 1.7.3 Research Aim and Objectives

The aim of this research project is to identify global cellular pathways or a 'genetic signature' that may be involved in the development of PD in a group of South African patients. This was investigated with the following objectives:

1. Select a subset of 20 PD patients and 20 controls from the Shared Roots parent study.
2. Perform mutation screening using MLPA and Sanger sequencing to identify pathogenic mutations in the patients.
3. Isolate RNA from whole blood and perform RNA-seq, as well as bioinformatic analysis of the data.
4. Replicate differentially expressed candidate genes using quantitative real-time PCR (qPCR).

# Chapter 2: Materials and Methods

2.1 Overview of the Methodology ..... 43

2.2 Study Participants ..... 44

2.2.1 Patients ..... 44

2.2.2 Controls ..... 44

2.3 Ethical considerations ..... 46

2.4 DNA Isolation and Methods ..... 46

2.4.1 Blood sample collection ..... 46

2.4.2 DNA Extraction ..... 46

2.4.3 MLPA Assay ..... 47

2.4.4 Primer Design for *PARK2* and *HBB* ..... 51

2.4.5 Sanger sequencing of the *PARK2* gene ..... 51

2.4.6 qPCR for verification of MLPA results ..... 54

2.5 RNA Isolation and Sequencing ..... 54

2.5.1 Blood Sample collection ..... 54

2.5.2 RNA Extraction ..... 55

2.5.3 In house RNA Quality and Integrity Analysis ..... 56

2.5.4 RNA-Seq Analysis ..... 57

*2.5.4.3 RNA-Seq on an Illumina HiSeq* ..... 58

2.6 Bioinformatic Data Analysis ..... 58

2.6.1 Base Calling ..... 59

2.6.2 Pre-alignment QC ..... 59

2.6.3 Alignment of reads to a reference ..... 59

2.6.4 Quantification of the transcriptome ..... 60

2.6.5 Differential gene expression ..... 60

2.6.6 Hierarchical clustering ..... 61

2.6.7 Pathway Analysis ..... 62

2.7 Gene Prioritisation ..... 62

2.8 Verification using qPCR ..... 63

2.8.1 PCR Primers ..... 63

2.8.2 RNA isolation ..... 64

2.8.3 cDNA Synthesis ..... 64

2.8.4 Standard real time PCR ..... 64

2.8.5 Data Analysis ..... 65

## 2.1 Overview of the Methodology

An overview of the methodology used in the present study is presented in Figure 2.1.

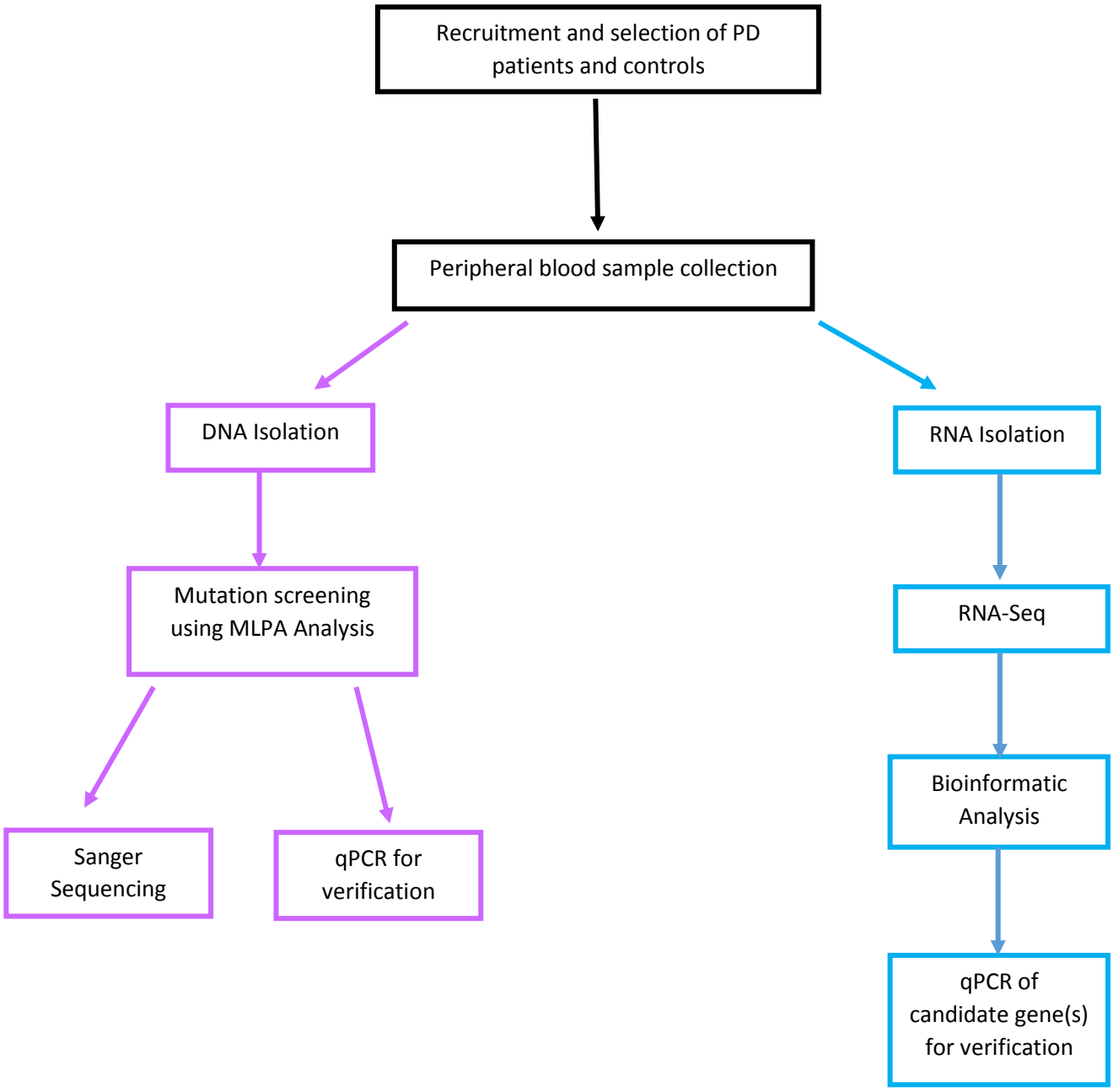



**Figure 2.1: Overview of methodology used for this study.** MLPA, multiplex ligation-dependant probe amplification; qPCR, quantitative real-time polymerase chain reaction

## 2.2 Study Participants

All relevant information on the 40 study participants is summarized in Table 2.1.

### 2.2.1 Patients

South African PD patients were recruited from the Movement disorders clinic at Tygerberg Hospital in Cape Town, South Africa. Inclusion criteria included patients who were adults (aged 18 years or older) and with a clinical diagnosis of PD for ≥4 years to exclude the likelihood of a misdiagnosis. Patients were diagnosed with PD according to the UK Brain Bank criteria (Hughes et al., 1992). They were excluded if they had significant head injury, had exposure to neuroleptics, or presented with atypical forms of PD. No patients had a family history of PD. The level of disability in patients was determined by Hoehn and Yahr scale (Hoehn and Yahr, 1967), and Part II and III of the Unified Parkinson's Disease Rating Scale (UPDRS) (Fahn et al., 1987). Cognitive function was assessed using the Montreal Cognitive Assessment (Nasreddine et al., 2005). Non motor function was assessed with the Non-Motor Symptom Assessment Scale for PD (Chaudhuri et al., 2007). Depression was assessed using the Hamilton Depression Scale (HAMD-D) (Hamilton, 1960). A total of 20 patients were selected for this study.

### 2.2.2 Controls

Twenty healthy controls were recruited through community newspaper advertisements and flyers and were matched to patients on age, ethnicity, and socioeconomic status. Table 2.2 summarises the demographic information of all PD patients and controls used in this study. There was no significant age difference between PD patients and controls (p-value= 0.076) showing that individuals were age matched.

**Table 2.1: Demographic and clinical characteristics of the 40 study participants.**

| Lab ID | Clinical Status | Age at sampling/ recruitment (yrs) | Gender | Ethnicity | HLOE | AAO of PD (yrs) | Disease Duration (yrs) | UPDRS Score (max=176 points) |
|---|---|---|---|---|---|---|---|---|
| **SR053** | PD | 69 | M | MA | Grade 10 | 63 | 6 | 61 |
| **SR033** | PD | 63 | M | MA | Grade 11 | 59 | 4 | 67 |
| **SR002** | PD | 37 | M | MA | Grade 11 | 32 | 5 | 87 |
| **SR005** | PD | 69 | F | MA | Grade 4 | 61 | 8 | 88 |
| **SR121** | PD | 67 | M | MA | Grade 10 | 52 | 15 | 90 |
| **SR011** | PD | 61 | F | MA | Grade 9 | 53 | 8 | 52 |
| **SR046** | PD | 53 | F | MA | Grade 7 | 43 | 10 | 102 |
| **SR023** | PD | 74 | M | MA | Grade 10 | 70 | 4 | 41 |
| **SR009** | PD | 61 | F | MA | Grade 12 | 49 | 12 | 55 |
| **SR056** | PD | 59 | M | MA | Grade 10 | 38 | 21 | 50 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **SR070** | PD | 62 | M | MA | Grade 10 | 55 | 7 | 125 |
| **SR041** | PD | 68 | F | MA | Grade 6 | 61 | 7 | 51 |
| **SR039** | PD | 55 | F | MA | Grade 8 | 49 | 6 | 77 |
| **SR014** | PD | 62 | M | MA | Grade 7 | 53 | 9 | 52 |
| **SR042** | PD | 43 | M | MA | Grade 10 | 39 | 4 | 49 |
| **SR067** | PD | 46 | F | MA | Grade 9 | 39 | 7 | - |
| **SR012** | PD | 54 | M | MA | Grade 10 | 49 | 5 | 82 |
| **SR007** | PD | 56 | M | MA | Post school diploma | 52 | 4 | 56 |
| **SR010** | PD | 75 | F | MA | Grade 10 | 62 | 13 | 72 |
| **SR003** | PD | 62 | M | MA | Grade 10 | 56 | 6 | 58 |
| **SR013** | HC | 54 | F | MA | Grade 7 | - | - | - |
| **SR098** | HC | 52 | F | MA | Grade 9 | - | - | - |
| **SR050** | HC | 55 | F | MA | Grade 8 | - | - | - |
| **SR079** | HC | 51 | F | MA | Grade 9 | - | - | - |
| **SR055** | HC | 59 | F | MA | Grade 9 | - | - | - |
| **SR103** | HC | 66 | F | MA | Grade 10 | - | - | - |
| **SR006** | HC | 61 | F | MA | Grade 8 | - | - | - |
| **SR089** | HC | 62 | F | MA | Grade 6 | - | - | - |
| **SR082** | HC | 62 | F | MA | Grade 6 | - | - | - |
| **SR109** | HC | 56 | F | MA | Grade 9 | - | - | - |
| **SR119** | HC | 54 | F | MA | Grade 5 | - | - | - |
| **SR099** | HC | 60 | F | MA | Grade 9 | - | - | - |
| **SR111** | HC | 57 | F | MA | Grade 9 | - | - | - |
| **SR080** | HC | 59 | F | MA | Grade 8 | - | - | - |
| **SR093** | HC | 55 | F | MA | Grade 7 | - | - | - |
| **SR088** | HC | 61 | F | MA | Grade 10 | - | - | - |
| **SR092** | HC | 59 | F | MA | Grade 8 | - | - | - |
| **SR054** | HC | 50 | F | MA | Grade 10 | - | - | - |
| **SR086** | HC | 59 | F | MA | Grade 8 | - | - | - |
| **SR027** | HC | 51 | F | MA | Post school diploma | - | - | - |

PD, Parkinson's disease; HC, Healthy Control; M, Male; F,Female; MA, Mixed ancestry; HLOE, Highest level of education AAO, Age-at-onset; yrs, years; UPDRS, Unified Parkinson's disease rating scale

**Table 2.2: Comparison of demographic information of the all study participants**

| | **PD Patients** | **Controls** | **p-value** | **OR [95% CI]** |
|---|---|---|---|---|
| **Age, mean ± SD** | 59.80 ± 9.89 | 57.15 ± 4.36 | 0.076 | 1.11 |
| **AAO ± SD** | 51.8 ± 9.79 | - | - | - |
| **Disease duration± SD** | 8.05 ± 4.36 | - | - | - |
| **Gender** | 40% female<br>60% male | 100% female | - | - |
| **Ethnicity** | Mixed ancestry | Mixed ancestry | - | - |

AAO, average age of onset; p-value, calculated probability; OR, odds ratio; CI, confidence interval

## 2.3 Ethical considerations

Written informed consent was obtained from all study participants, an example of the consent form is presented in Appendix VI. Study approval was obtained from the Health Research Ethics Committee (HREC), Stellenbosch University, Cape Town (Approval number: N13/08/115). Ethical approval was renewed annually. All study participant information was stored on Research Electronic Data Capture (REDCap) (https://projectredcap.org/), a secure database for data capture of participant information.

## 2.4 DNA Isolation and Methods

### 2.4.1 Blood sample collection

Peripheral blood for DNA isolation was drawn from all study participants. Ten to 15 ml of blood was drawn by means of venous puncture and collected in two 10 ml ethylene-diamine-tetra-acetic acid (EDTA) tubes by a trained nurse and was delivered to the MAGIC Laboratory in the Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University. Blood samples were stored at 4 °C for no longer than four days before DNA was extracted.

### 2.4.2 DNA Extraction

DNA was isolated from the blood using an in-house phenol/chloroform method by Miss Rolanda Londt (lab technician of SHARED Roots study). Blood was pooled from both 10 ml EDTA tubes (per participant) into a 50 ml centrifuge tube. The tube was then filled with 20 ml ice-cold lysis buffer (Appendix VIII) and gently inverted several times. The tube was then incubated on ice for 10 minutes and subsequently centrifuged at 3000 rpm for 10 minutes at 4˚C in a Beckman model TJ-6 centrifuge (Beckman Coulter Inc., Scotland, UK). The supernatant was discarded and the pellet resuspended in 20 ml ice cold lysis buffer which was then incubated for 10 minutes and centrifuged at 3000 rpm at 4˚C. The supernatant was discarded and the pellet resuspended in 900 µl Sodium-EDTA and 100 µl 10% SDS.

A total volume of 100 µl Proteinase K (10mg/ml) was added to each 50 ml tube and the pellet resuspended by vortexing and incubated overnight at 37˚C. 2 ml distilled water and 500 µl of sodium acetate (3M) to each 50 ml tube and mixed by vortexing. After this step, 2.5 ml phenol-chloroform solution (1:1, v/v) was added to each 50 ml tube and lids closed tightly. Tubes were placed on a Voss rotator (Voss of Maldon, England) at a 45˚ angle and shaken at 140 rpm for 10 minutes to obtain an emulsion. The mixture was then poured into a 10 ml Corex glass tube and spun at 8000 rpm for 12 minutes at 4˚C. The upper aqueous phase was gently transferred to a clean Corex glass tube using a Pasteur pipette. 2.5 ml Chloroform-Octanol solution (24:1, v/v) was added to each Corex glass tube containing the aqueous phase and sealed with silicone rubber tops and mixed by inversion until the

homogenate turned 'milky'. The rubber tops were removed and Corex tubes centrifuged in a Sorvall RC-5B refrigerated super-speed centrifuge (rotor SS 34, Dupont Instruments) at 8000 rpm for 12 minutes at 4˚C.

The upper aqueous phase was then transferred to a 15 ml plastic Greiner tube. DNA was precipitated by adding 7 ml of 95% ice-cold ethanol, tubes were sealed and inverted until DNA precipitates out of solution and was visible as a thread-like clot. The DNA precipitate was transferred with a pipette into a clean 1.5 ml Eppendorf microfuge tube. To each Eppendorf tube, 1 ml of 70% ethanol was added and inverted several times to wash the DNA and left at room temperature for 1 minute. After this step the Eppendorf tube was centrifuged at 14000 rpm for 3 minutes and ethanol removed by decanting. After removal of most of the ethanol, the pellet was air-dried at room temperature for 10 minutes.

The dried DNA was resuspended in 300 µl TE solution and incubated overnight at 37˚C and subsequently for 1 hour at 50˚C. Eppendorf microfuge tubes were tightly sealed and placed in a 50 ml centrifuge tube which were then placed on Voss rotator and spun at 30 rpm for 3 days. DNA was then quantified using the Nanodrop™ 2000 (*Thermo Scientific, MA, USA)* at 260 nm. DNA quality is a ratio of absorbance (A260/A280), where a ratio of 1.8 indicates good quality DNA. A DNA ratio lower than 1.8 indicates protein contamination whereas a higher value indicates RNA contamination. All DNA was stored at -80˚C until required for Multiplex ligation-dependent probe amplification (MLPA) analysis and Sanger sequencing.

### 2.4.3 MLPA Assay

MLPA is a quantitative PCR-based laboratory technique which can detect genomic rearrangements and exonic deletions or duplications.

MLPA was performed on all patient samples using two commercially available kits: SALSA P051-D1 and P052-D1 Parkinson MLPA kits (MRC-Holland, Amsterdam, the Netherlands) according to the manufacturer's instructions. The P051-D1 and P052-D1 kits contain specifically designed probes for exons of PD-causing genes including: α-synuclein (*SNCA)*, *PARK2* (*Parkin*), *UCHL1, PINK1, PARK7, LRRK2,* GTP Cyclohydrolase 1 *(GCH1)* and *ATP13A2*. The P051-D1 Parkinson probemix-1 and P052-D1 Parkinson probemix-2 each contain 50 MLPA probes including two probes specific for the *SNCA* A30P and *LRRK2 G2019S* mutations. Amplification products are between 130 nt and 500 nt. In addition, each probemix

contains nine control fragments which includes four DNA Quantity fragments (Q-fragments) at 64, 70, 76 and 82 nt, three DNA Denaturation control fragments (D-fragments) at 88, 92 and 96 nt, one X-fragment at 100 nt and one Y-fragment at 105 nt. MLPA can therefore be used for gender detection.

Probes present in the P051-D1 Parkinson probemix-1 include: *PARK7* (9 probes), *ATP13A2* (2 probes), *PINK1* (8 probes), *SNCA* (7 probes, including one for the *SNCA* A30P mutation), *PARK2* (12 probes), *LRRK2* (1 probe for the G2019S mutation) and 10 reference probes.

The P052-D1 Parkinson probemix-2 includes the following probes: *ATP13A2* (2 probes), *UCHL1* (8 probes), *PARK2* (12 probes), *PACRG* (1 probe), *LRRK2* (8 probes including G2019S mutation), *GCH1* (6 probes), *CAV1* (1 probe), *CAV2* (1 probe) and 11 reference probes.

### *Principle of MLPA*

MLPA involves the amplification of probes to detect specific DNA sequences of approximately 60 nt in length and specific exonic deletions and duplications in genes. MLPA consists of four distinct steps namely; denaturation and hybridization, ligation, polymerase chain reaction and fragment analysis (Figure 2.2). Each probe in the MLPA mix contains two oligonucleotides which recognize a specific sequence on the DNA. One of the oligonucleotides also contains a stuffer sequence which is different for each probe making the oligonucleotide longer in length. The stuffer sequence has a unique length for each probe allowing separation according to a unique size. One oligonucleotide contains the sequence for the forward primer and the second oligonucleotide for the reverse primer. Only when both oligonucleotides are hybridized to the correct template and are adjacent to each other, will ligation occur. Only ligated probes will be amplified during the PCR reaction. Each PCR product has a unique length and amplification products can therefore be separated according to size using capillary electrophoresis.

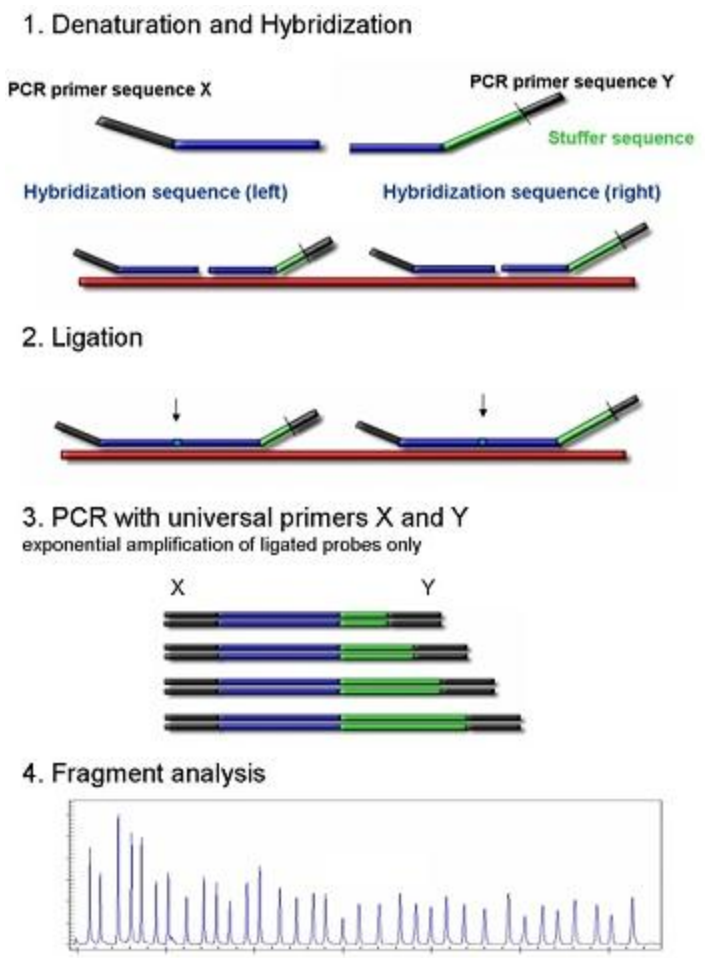



**Figure 2.2: Multiplex ligation-dependent probe amplification (MLPA) assay principle and the various steps involved in the experiment.**

A total quantity of 100 ng genomic DNA was diluted in 0.1M TE buffer to 5 µl and denatured for 5 minutes at 98°C and hybridized with MLPA buffer and MLPA probe sets P051 and P052 (in independent reactions for each probe set) at 60°C for 16 hours using the 2720 Thermal Cycler (Applied Biosystems, Foster City, USA). Ligase 65 master mix (quantity provided in Table 2.3) was added to each sample at 54°C and ligation was then performed for 15 minutes at 54°C following a heat inactivation step of Ligase-65 enzyme for 5 minutes at 98°C. A SALSA PCR primer mix (Table 2.3) was then prepared and added to each sample and amplification was performed for 35 cycles of 30 seconds at 95°C, 30 seconds at 60°C, 60 seconds at 72°C and incubation for 20 minutes at 72°C and a cool down at 15°C.

**Table 2.3: MLPA Master-mix reagents and volumes used for an MLPA assay.**

| Master Mix | Volume per sample (µl) |
|---|---|
| **Hybridization master mix:** | |
| MLPA Buffer | 1.5 |
| Probemix (P051-D1 or P052-D1) | 1.5 |
| **Ligase-65 master mix:** | |
| dH2O | 25 |
| Ligase Buffer A | 3 |
| Ligase Buffer B | 3 |
| Ligase-65 enzyme | 1 |
| **SALSA PCR master mix:** | |
| dH2O | 7.5 |
| SALSA PCR primer mix | 2 |
| SALSA polymerase | 0.5 |

dH2O, distilled water

The PCR products were electrophoresed at the Central Analytical Facility (CAF) (Stellenbosch University, Stellenbosch, South Africa) using capillary electrophoresis on an ABI 310 genetic analyzer. Coffalyser.net version 140721.1958 (MRC, Holland), a freely-available MLPA analysis software tool (https://coffalyser.wordpress.com/downloads/)) was used for DNA fragment analysis and comparative analysis to determine peak areas of PCR products and to calculate copy number. Data were normalized by dividing each probe's peak area by the average peak area of the reference peaks. The results from each sample were compared to a dosage quotient to obtain the copy number status (Table 2.4).

**Table 2.4: Copy number and dosage quotient values using Coffalyser.net software**

| Copy number status | Dosage quotient |
|---|---|
| Normal | 0.80-1.20 |
| Homozygous deletion | 0.00 |
| Heterozygous deletion | 0.40-0.65 |
| Heterozygous duplication | 1.30-1.65 |
| Heterozygous triplication/ homozygous duplication | 1.75-2.15 |
| Ambiguous | All other values |

### 2.4.4 Primer Design for *PARK2* and *HBB*

Oligonucleotide primers required for PCR were designed using Primer 3 software version 4.0.0 (http://primer3.ut.ee) (Koressaar and Remm, 2007) using published sequence data from the National center for biotechnology information (NCBI) Genbank database (http://www.ncbi.nlm.nih.gov/Entrez) or Ensembl Genome Browser database (http://www.ensembl.org). IDT OligoAnalyzer® software version 3.1 (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer) was utilized to test primer-primer complementarity, self-complementarity and melting temperature compatibility. Specificity of primer binding sites was also verified using Basic Local Alignment Search Tool (BLAST) (http://www.ncbi.nlm.nih.gov/BLAST). Forward (F) and reverse (R) primers were synthesized for the gene of interest, *PARK2* and a reference gene namely haemoglobin subunit beta (*HBB)* using standard phosphoramidite chemistry at Integrated DNA Technologies (IDT, Iowa, USA). Primer sequences are provided in Table 2.5 (*PARK2* gene) and Table 2.6 (*HBB* gene). The HBB sequence and location of the forward and reverse primers (Appendix VII) and the expected PCR product sizes are shown in Figure 2.3.

### 2.4.5 Sanger sequencing of the *PARK2* gene

A total reaction mixture of 25 µl was set up for each sample consisting of the following reagents: 20 µM of forward and reverse primers, 1.5 mM $MgCl_2$ (Bioline, London, UK), 2.5 mM deoxynucleotides (dNTPs) (Promega, Madison, Wisconsin, USA), 1x Ammonium ($NH_4$) buffer (Bioline, London, UK); 0.25 units BIOTAQ DNA Polymerase (Bioline, London, UK). Addition of 10% dimethyl sulfoxide (DMSO) and Formamide were required for *PARK2* exons 1 and 2, respectively. DMSO and Formamide are additives which increase amplification in a PCR reaction. The volume of nuclease free water was adjusted accordingly.

All 12 exons of *PARK2* were PCR amplified using the 2720 Thermal Cycler (Applied Biosystems, CA, USA) with the following cycling conditions: an initial step at 94˚C for 5 minutes, followed by 35 cycles of denaturation at 94˚C for 15 seconds, varying annealing temperatures specific for each exon (Table 2.5) for 30 seconds and elongation at 72˚C for 30 seconds. Followed by two holding steps: 72˚C for 7 minutes and 25˚C for 1 minute. Gel electrophoresis was performed and the PCR products were separated on a 1% sodium borate (SB) agarose gel, which contained Ethidium bromide (EtBr) used for visualization, along with a size ladder, the Universal ladder (Kapa Biosystems, Cape Town, South Africa) at 120 volts for one hour (Appendix VIII). Each run contained two positive wild type (WT) control samples and a negative control. PCR products were visualized using the GeneSnap Transilluminator (Model 6.07, Synoptics Ltd).

To ensure clean and readable DNA sequencing results, PCR products were cleaned by adding 0.05 units Shrimp alkaline phosphatase (Sap I) (Cleveland, Ohio, USA), which removes the dNTPs, and 0.05 units Exonuclease (Exo I) (Promega, USA), which removes leftover primers, to 8 µl PCR product. The reaction mixture was then incubated at 37°C for 15 minutes, followed by 85°C for 15 minutes in the 2720 Thermal Cycler (Applied Biosystems, CA, USA) to deactivate the enzymes. The concentration of cleaned PCR products was then determined using the *Nanodrop™ 2000 (*Thermo Scientific, MA, USA). Samples were diluted accordingly to obtain a final concentration of 30 ng/µl. The F and R *PARK2* primers were diluted to a final concentration of 1.1 µM for the sequencing rection.

Sequencing was performed at CAF, using Sanger Sequencing. The BigDye Terminator Sequence Ready Reaction kit version 3.1 (Applied Biosystems, CA, USA) was used. Reactions were electrophoresed and analysed on a 3130 x1 Genetic Analyser (Applied Biosystems, CA, USA) and sequencing results were analysed using BioEdit version 7.0.1 (Hall 1999).

**Table 2.5: Primers designed for the exonic sequences of the *PARK2* gene**

| Region | Primer Sequence (5'-3') | Tm (˚C) | % GC | PCR Ta (˚C) | PCR Conditions | Size of PCR product (bp) |
|---|---|---|---|---|---|---|
| Exon 1 | For: gaa cta cga ctc cca gca g<br>Rev: ccc gtc att gac agt tgg | 60<br>56 | 58<br>56 | 55 | DMSO<br>35 cycles | 300 |
| Exon 2 | For: cac cat tta agg gct tcg ag<br>Rev: tca ggc atg aat gtc aga ttg | 60<br>60 | 50<br>43 | 55 | Formamide<br>35 cycles | 313 |
| Exon 3 | For: tct cgc att tca tgt ttg aca<br>Rev: gca gac tgc act aaa caa aca | 58<br>62 | 39<br>43 | 55 | No additives<br>35 cycles | 364 |
| Exon 4 | For: gct ttt aaa gag ttt ctt gtc<br>Rev: ttt ctt ttc aaa gac ggg tga | 56<br>58 | 33<br>38 | 55 | No additives<br>35 cycles | 299 |
| Exon 5 | For: gga aac atg tct taa gga gt<br>Rev: ttc ctg gca aac agt gaa ga | 56<br>58 | 40<br>45 | 55 | No additives<br>35 cycles | 223 |
| Exon 6 | For: cca aag aga ttg ttt act gtg<br>Rev: ggg gga gtg atg cta ttt tt | 58<br>58 | 38<br>45 | 55 | No additives<br>35 cycles | 276 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Exon 7 | For: cct cca gga tta cag aaa ttg<br>Rev: gtt ctt ctg ttc ttc att agc | 60<br>58 | 43<br>38 | 55 | No additives<br>35 cycles | 280 |
| Exon 8 | For: ggc aac act ggc agt tga ta<br>Rev: ggg gag ccc aaa ctg tct | 60<br>58 | 50<br>61 | 55 | No additives<br>35 cycles | 232 |
| Exon 9 | For: tcc cat gca ctg tag ctc ct<br>Rev: cca gcc cat gtg caa aag c | 62<br>60 | 55<br>58 | 55 | No additives<br>35 cycles | 297 |
| Exon 10 | For: caa gcc aga gga atg aat at<br>Rev: gga act ctc cat gac ctc ca | 56<br>62 | 40<br>55 | 53 | No additives<br>35 cycles | 272 |
| Exon 11 | For: ccg acg tac agg gaa cat aaa<br>Rev: ggc acc ttc aga cag cat ct | 62<br>62 | 48<br>55 | 55 | No additives<br>35 cycles | 300 |
| Exon 12 | For: tct agg cta gcg tgc tgg tt<br>Rev: gcg tgt gtg tgt gtg ttt ga | 62<br>60 | 55<br>50 | 55 | No additives<br>35 cycles | 296 |

Tm, melting temperature; Ta, annealing temperature; bp, base pairs

**Table 2.6: Primer sequences for *HBB* reference gene**

| Name of Primer | Primer Sequence (5'-3') | Tm (˚C) | %GC | Ta (˚C) | Size of PCR product (bp) |
|---|---|---|---|---|---|
| HBB-F | ACACAACTGTGTTCACTAGC | 58 | 45 | 55 | Genomic: 377<br>mRNA: 247 |
| HBB-R | ACCGAGCACTTTCTTGCCAT | 60 | 50 | | |

Tm, melting temperature; Ta, annealing temperature; bp, base pairs

```
> 377bp on genomic template
ACACAAC TGTGTTCACT AGCAACCTCA AACAGACACC ATGGTGCATC

    TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG
                                                                    TTGGTGGTGA
GGCCCTGGGC AGgttggtat caaggttaca agacaggttt aaggagacca
    atagaaactg ggcatgtgga gacagagaag actcttgggt ttctgatagg cactgactct

    ctctgcctat tggtctattt tcccaccctt agGCTGCTGG TGGTCTACCC TTGGACCCAG
    AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG
  GTGAAGGCTC ATGGCAAGAA AGTGCTCGGT
```

**Figure 2.3: PCR products of the *HBB* primer set.** Forward and reverse primer sequences are highlighted in yellow. The intron is 130 bp and therefore the product size of gDNA is 377 bp and 247 bp on cDNA (when the intron is spliced out)

### 2.4.6 qPCR for verification of MLPA results

Copy number variation (CNV) mutations found were verified using an independent technique namely quantitative real-time PCR (qPCR). A total reaction master mix of 20 µl was prepared for each primer which included: 20 µM Forward primer, 20 µM Reverse primer, 1x RNase free water, 1x FastStart Essential DNA Green Master Mix (Roche Diagnostics, Indianapolis, USA) and 100 ng genomic DNA (gDNA) template. Reaction volumes were transferred into a 96 well plate using the epMotion®5070f (Eppendorf, Hamburg, Germany). The 96 well plate was then sealed and centrifuged at 20˚C for 1 minute at 1500 rpm on the Eppendorf 5810 R centrifuge *(*Eppendorf, Hamburg, Germany).

Plates were run on the Lightcycler® 96 (Roche Diagnostics, Indianapolis, USA ) with the following cyclic conditions: an initial preincubation step at 95˚C for 600 seconds, followed by 45 cycles of denaturation at 95˚C for 15 seconds, annealing at 55˚C for 20 seconds and elongation at 72˚C for 20 seconds. This was followed by melting at 95˚C for 600 seconds, 65˚C for 60 seconds and 95˚C for 1 second. A final cooling step at 25˚C for 200 seconds was included. Upon completion the plate was removed and data stored and analysed using the Lightcycler® 96 software version 1.1 (Roche Diagnostics, Indianapolis, USA).

## 2.5. RNA Isolation and Sequencing

### 2.5.1 Blood Sample collection

Peripheral blood samples for RNA extraction were obtained by means of venous puncture from each PD patient and control at the same time as the bloods that were taken for DNA isolation. Blood was drawn and collected in 2.5 ml PAXgene blood RNA tubes (Qiagen, Hilden, Germany). Immediately after blood collection, the PAXgene tube was inverted 8-10 times. The PAXgene tubes with participants blood was kept at room temperature. A more detailed protocol on the procedure followed for blood collection in Paxgene tubes is supplied in Appendix IX. Blood samples were then delivered to the MAGIC Laboratory

in the Division of Molecular Biology and Human Genetics at Stellenbosch University within 2 hours of collection and were stored upright at 4˚C for 24 hours. Bloods were then transferred to the -20˚C freezer for 24 hours and thereafter stored at -80˚C where blood RNA is stable for up to 50 months (Paxgene blood RNA Kit Handbook, PreAnalytix).

### 2.5.2 RNA Extraction

For RNA isolation, the PAXgene blood RNA tubes were removed from -80˚C storage and incubated at room temperature for 2 hours to ensure complete lysis of the blood cells. Work surfaces and equipment were decontaminated using RNaseZap® (ThermoFisher Scientific, MA, USA). All pipette tips and tubes used were RNase free. All reagents used were supplied in the PAXgene™ Blood RNA Kit (Qiagen, Hilden, Germany). The centrifugations were carried out at 15˚C in the Eppendorf 5810 R centrifuge *(*Eppendorf, Hamburg, Germany). RNA was extracted from all PD patient and control samples using the manual protocol (PAXgene Blood RNA Kit Handbook, Version 2), as follows:

PAXgene Blood RNA tubes were centrifuged for 20 minutes at 4000 x g. The supernatant was then removed by decanting and 4 ml of RNase-free water was added to the pellet and tubes closed using a fresh secondary BD Hemogard closure. The pellet was thoroughly resuspended by pipetting and vortexing and tubes were then centrifuged for 10 minutes at 4000 x g. The supernatant was removed and discarded. The pellet was then resuspended in 350 µl resuspension buffer (BR1) and vortexed until the pellet was visibly dissolved. The entire sample was then pipetted into a 1.5 ml microcentrifuge tube (MCT) and 300 µl binding buffer and 40 µl proteinase K (PK) was added to each tube. The contents of the tube were mixed by vortexing and incubated on a shaker-incubator for 10 minutes at 55˚C at 1000 rpm.

The lysate was pipetted into a lilac coloured PAXgene Shredder spin column (PSC) placed in a 2 ml processing tube (PT) and centrifuged for 3 minutes at 15 000 x g. The supernatant was transferred to a 1.5 ml microcentrifuge tube without disturbing the pellet in the processing tube. 350µl 100% ethanol was added to the tubes and the contents of the tubes mixed by vortexing and centrifuged briefly (1-2 seconds). 700 µl of the sample was added into a red coloured PAXgene RNA spin column (PRC) and centrifuged for 1 minute at 15 000 x g. The spin column was placed in a new 2 ml processing tube and the old processing tube (containing the flow through) was discarded. The remaining sample in the microcentrifuge tube was then pipetted into the same PAXgene RNA spin column and centrifuged for 1 minute at 15 000 x g. The spin column was then placed in a new 2 ml processing tube and the old processing tube was discarded.

350 µl wash buffer 1 (BR3) was pipetted into the PRC and centrifuged for 1 minute at 15 000 x g. The spin column was then placed in a new 2 ml processing tube and the old processing tube was discarded.

Preparation of DNase I: lyophilized DNase I was carefully dissolved in 550 µl of DNase resuspension buffer (DRB) and mixed by inversion of the tube. For each sample, 10 µl DNase I stock solution and 70 µl DNA digestion buffer (RDD) was added to a 1.5 ml microcentrifuge tube, mixed by inversion and centrifuged briefly. The DNase I incubation mix was directly pipetted onto the PAXgene RNA spin column membrane and incubated at room temperature for 15 minutes.

350 µl wash buffer 1 (BR3) was pipetted into the PAXgene RNA spin column and centrifuged for 1 minute at 15 000 rpm. The spin column was then placed in a new 2 ml PT and the old PT was discarded. 500 µl wash buffer 2 (BR4) was pipetted into the PAXgene RNA spin column and centrifuged for 1 minute at 15 000 rpm. The old PT was discarded and the spin column placed in a new 2 ml PT. An additional 500 µl wash buffer 2 (BR4) was added to the PAXgene RNA spin column and centrifuged for 3 minutes at 15 000 rpm. The spin column was placed in a new 2 ml PT and the old tube discarded. The PAXgene RNA spin column was then centrifuged for 1 minute at 15000 rpm. The spin column was then placed in a 1.5 ml microcentrifuge tube and the old processing tube discarded. 40 µl elution buffer (BR5) was added to the PAXgene RNA spin column and centrifuged for 1 minute at 15000 rpm to elute the RNA. The elution step was repeated using an additional 40 µl elution buffer (BR5) and the same microcentrifuge tube. The eluate was incubated at 65˚C for 5 minutes in the shaker-incubator without shaking and then chilled immediately on ice. The RNA quality and integrity was determined for all RNA samples using the Agilent Bioanalyser (Agilent Technologies, Santa Clara, United States) and the stock RNA was stored at -80˚C.

### 2.5.3 In house RNA Quality and Integrity Analysis

#### *2.5.3.1 RNA Quality and Integrity Analysis using the Agilent Bioanalyzer 6000 Nano Kit*

RNA quality and integrity was determined using the Agilent Bioanalyzer Nano 6000 kit housed at CAF (Stellenbosch University, Stellenbosch, South Africa) on the Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, United States). The procedure was performed according to the Agilent RNA 6000 Nano Assay Protocol (Appendix X).

The RNA Integrity Number (RIN) was calculated algorithmically with the RIN software algorithm which includes the 28s/18s ratio, the region before the peaks, signal areas and intensities. The RIN number calculation uses microcapillary electrophoretic RNA separation to determine the RNA integrity of samples containing total RNA (Schroeder et al., 2006). RIN values range from 1 to 10, where a RIN value

of 10 represents fully intact RNA and RIN value of 1 represents totally degraded RNA. A minimum RIN value of 7 was required for samples to be sequenced.

#### *2.5.3.2 In-house RNA Quality Control*

As an additional check that the extracted RNA was of good quality and with minimal gDNA contamination, an additional in-house quality control (QC) procedure was performed. For this, template RNA from one control sample was converted to cDNA using the QuantiNova™ Reverse Transcription Kit (Qiagen, Hilden, Germany). A total reaction mixture of 14 µl was set up for each sample consisting of 1x DNA wipeout buffer, RNase-free water added to 750 ng template RNA and incubated at 45˚C for 2 minutes using the 2720 Thermal Cycler (Applied Biosystems, CA, USA). Samples were then immediately placed on ice. Reverse transcription master mix was prepared and added to each sample consisting of the following reagents: Reverse transcription enzyme and reverse transcription mix (which contains $Mg^{2+}$ and dNTPs) at a final reaction volume of 19µl. Samples were incubated at 25˚C for 3 minutes for annealing, reverse transcription at 45˚C for 10 minutes followed by 85˚C for 5 minutes to inactivate the Reverse Transcriptase Enzyme (Appendix XI). All samples were placed on ice and subsequently stored at -80˚C if not immediately required for further experiments.

A total PCR reaction mixture of 25 µl was set up for each cDNA and gDNA template consisting of the following reagents: 20 µM of forward and reverse primers for hemoglobin subunit beta (HBB), 3.0 mM $MgCl_2$ (Bioline, London, UK), 2.5 mM dNTPs (Promega, Madison, Wisconsin, USA), 1x Ammonium ($NH_4$) buffer (Bioline, London, UK), 0.25 units BIOTAQ DNA Polymerase (Bioline, London, UK) and 1µl cDNA or gDNA. PCR amplification was performed using the 2720 Thermal Cycler (Applied Biosystems, CA, USA) with the following cycling conditions: denaturation of 94˚C for 5 minutes followed by 35 cycles of 94˚C for 30 seconds, 59˚C for 30 seconds, 72˚C for 1 minute and elongation at 72˚C for 7 minutes. The PCR products were electrophoresed on a 1% sodium borate (SB) agarose gel, which contained Ethidium bromide (EtBr), along with a size ladder, the Universal ladder (KAPA Biosystems, Cape Town, South Africa) at 120 volts for one hour. PCR products were visualized using the GeneSnap Transilluminator (Model 6.07, Synoptics Ltd).

### 2.5.4 RNA-Seq Analysis

#### *2.5.4.1 Preparation of RNA samples for sequencing:*

RNA had to be of good quality and at a high concentration to ensure effective RNA Sequencing. A total of 550 ng RNA was required. RNA was aliquoted into RNase free tubes according to calculated volumes

(Appendix XII, Table 6) and exported at -80˚C using dry ice via courier to NXT-Dx (Gent, Belgium) for RNA-seq analysis.

#### *2.5.4.2 cDNA library preparation performed by NXT-Dx*

The TruSeq Stranded Total RNA Library Prep Kit (Illumina, San Diego, USA) was used for library preparation as per standard protocols, in brief sample libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina, San Diego, USA), which included the use of Illumina in-line control spike-in transcripts. Library preparation used 500 ng of RNA in 50 µl of nuclease-free water, which was subjected to poly (A)+ purification using oligo-dT magnetic beads. After washing and elution, the polyadenylated RNA was fragmented to a median size of ˜150 bp and then used as a template for reverse transcription using random primers. The resulting single-stranded cDNA was converted to double-stranded cDNA, the ends were repaired to create blunt ends, a single (A) residue was added to the 3' ends to create (A)-tailed molecules. Illumina indexed sequencing adaptors were then ligated to the (A)-tailed double-stranded cDNA. A single index was used for each sample. The adaptor-ligand cDNA was then subjected to PCR amplification for 15 cycles. The final library product was purified using AMPure beads (Beckman Coulter Inc., Scotland, UK), quantified by qPCR (Kapa Biosystems, Cape Town, South Africa), and the size distribution assessed using the Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, United States). Following quantitation, an aliquot of the library was normalized to 2 nM concentration and equal volumes of specific libraries were mixed to create multiplexed pools in preparation for sequencing.

#### *2.5.4.3 RNA-Seq on an Illumina HiSeq*

cDNA libraries were sequenced (paired end, 50 bp) using the Illumina HiSeq 4000 (Illumina Inc, San Diego, USA) by NXT-Dx (Gent, Belgium). The Illumina HiSeq systems utilize sequencing by synthesis and patterned flow cell technology which contain billions of nanowell substrates at fixed locations across both surfaces of the flow cell. The structured organization allows detection of significant increases in sequencing reads and total read output. HiSeq 4000 v3.3.20 Software (Illumina Inc, San Diego, USA) was utilized to generate raw sequencing reads.

## 2.6 Bioinformatic Data Analysis

Data analysis was performed using Partek® Flow® Software build 4.0.15 (Partek Incorporated, St Louis, MO, USA) by CAF (Stellenbosch University, Stellenbosch, South Africa). Raw read files were obtained from NXT-Dx (Gent, Belgium). A summary of the various steps in the analysis pipeline using Partek® Flow® is depicted in Figure 2.4.

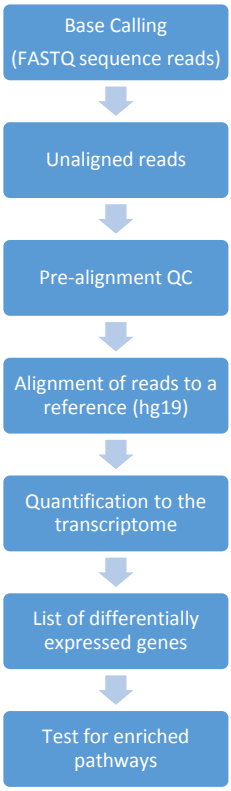



**Figure 2.4: Bioinformatic analysis pipeline of RNA-Seq data using Partek® Flow®.** hg19, also known as GRCh37, QC, quality control

### 2.6.1 Base Calling

FASTQ sequence reads were generated using the Illumina Casava Pipeline 1.8.2. Each sample had two fastq files.

### 2.6.2 Pre-alignment Quality Control

Quality assessment was performed using FASTQC quality control tool version 0.11.4. FastQ files were imported into Partek® Flow® (Partek Incorporated, St Louis, MO, USA). FASTQC performed an analysis of raw sequencing reads and generated reports. These reports included per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences and Kmer content, which detects duplicated sequences.

### 2.6.3 Alignment of reads to a reference

Reads were mapped using Spliced Transcripts Alignment to a Reference (STAR) aligner v2.3.1j. The parameters used are provided in Appendix XIII, Table 7.

### 2.6.4 Quantification of the transcriptome

BAM files generated were mapped on the human reference genome build hg19 (GENCODE) using Partek expectation-maximization (E/M) algorithm. Table 2.7 shows the default input parameters used.

**Table 2.7: Partek E/M default input parameters.**

| Parameter | Value |
|---|---|
| Strict paired-end compatibility | True |
| Require junction reads to match introns | True |
| Strand specificity | Auto-detect (Default: No) |
| Minimum read overlap with feature | 80 |
| Report unexplained regions | False |
| Include BAM files in output project file | False |

BAM, binary alignment/map; E/M, expectation-maximization

### 2.6.5 Differential gene expression

Expression analysis was performed using gene-specific analysis (GSA) a statistical modeling approach to test for differential expression of genes or transcripts in Partek® Flow®. Raw read counts and RPKM values were calculated for each annotated gene and transcript. Input parameters for GSA are shown in Table 2.8, non-default parameters are displayed in red. An unfiltered candidate gene list was generated which was further filtered using parameters listed in Table 2.9, which produced a filtered candidate gene list.

**Table 2.8: GSA input parameters.**

| Parameter | Value |
|---|---|
| Run analysis on | Gene level |
| Filter type | Minimum coverage--- 120 |
| Normalization offset | 1.0E-4 |
| Normalization type | RPKM |
| FDR step-up | True |
| Storey q-value | False |
| Min error degrees of freedom | 0 |

| | |
|---|---|
| Model selection criterion | AICc |
| Enable multimodel approach | Yes |
| Use only reliable estimation results | Yes |
| p-value type | Wald |
| Normal | True |
| Lognormal | True |
| Poisson | True |
| Negative binomial | True |

*Red colour indicates parameters that were changed from the default parameters. FDR, false discovery rate; GSA, gene specific analysis; RPKM, reads per kilobase of transcript per million mapped reads.

**Table 2.9: Filtered gene list parameters**

| Parameter | Value |
|---|---|
| p-value | < 0.01 |
| False Discovery Rate | < 0.05 |
| Fold Change | ≤ -2 or ≥ +2 |

### 2.6.6 Hierarchical clustering

Hierarchical clustering was performed using on the filtered gene list in the Partek® Flow® Software build 4.0.15 (Partek Incorporated, St Louis, MO, USA). Clustering was based on combining two of the most similar clusters based on similarity. No experimental variables were taken into account such as disease group.

A dendogram was produced and the expression of each gene was standardized to a mean of 0 and standard deviation of one. Unchanged genes had a value of 0 and were grey coloured. Upregulated genes had a positive value and were displayed as red whereas down-regulated genes were displayed as green with a negative value. The heat map showed the standardized gene expression level of each gene in each sample.

### 2.6.7 Pathway Analysis

#### *2.6.7.1 Gene ontology enrichment*

Genes were classified with the Gene ontology tool (www.geneontology.org) Enrichment analysis was performed using the candidate gene list using the PANTHER classification system. A p-value of ≤ 0.05 was the parameter for the analysis biological process, molecular function and cellular component options.

#### *2.6.7.2 Functional annotation clustering*

Functional annotation clustering was performed using the database for annotation, visualization and integrated discovery (DAVID) Bioinformatics Database (https://david.ncifcrf.gov/) to identify enriched functional gene related groups and enriched biological themes

#### *2.6.7.3 Ingenuity pathway analysis (IPA)*

Genes with significant differential expression were also subjected to Ingenuity pathway analysis (IPA) using IPA 7.1 software (IPA, Ingenuity Systems, www.ingenuity.com) to identify significant dysregulated canonical pathways and to generate gene networks. The IPA library identified most significant pathways within the filtered gene list. The p-value was calculated using Fisher's exact test to calculate the likelihood that the association between the genes in the dataset and canonical pathways is explained by chance alone. Gene networks were generated by overlaying the filtered genes in our dataset onto global molecular networks and algorithmically generated based on their connectivity.

## 2.7 Gene Prioritisation

The filtered candidate gene list with a ≤ -2 or ≥ +2 fold change, significant p-value of < 0.01 and false discovery rate (FDR) score of < 0.05 were further subjected to a literature search and pathway analysis to identify plausible candidate genes of significant biological relevance to PD. To assess biological relevance of the candidate genes, a literature search of previous gene expression studies on PD was performed. Gene expression studies in blood and brain were included and tables were compiled for these studies which included 4,199 candidate genes from blood studies and 12, 117 candidate genes from brain studies. These tables are available in Appendix I and II, Tables 1 and 2. The candidate gene list generated from this study was compared to these compiled tables. Genes were prioritized if they were differentially expressed in the same direction in a minimum of one blood, one brain study and the candidate gene list in the present study.

The National Center for Biotechnology Information (NCBI) (http://www ncbi.nlm nih.gov/RefSeq/), Online Mendelian Inheritance in Man (OMIM) (http://www.omim.org/) and GeneCards® (http://www.genecards.org/) databases were searched for identification of the function of these biologically relevant candidate genes.

## 2.8 Verification using qPCR

Quantitative real-time PCR (qPCR) is an alternative method to perform gene expression analysis. Plausible candidate genes identified from RNA-Seq were verified using this technique. All qPCR experiments were performed in accordance with the Minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines (Bustin et al., 2009).

### 2.8.1 PCR Primers

Commercially available primers namely the QuantiTect®Primer assays (Qiagen, Hilden, Germany) were purchased for selected genes, which included one target gene and three reference genes. CCAAT/enhancer binding protein alpha *(CEBPA)* was selected as the target gene and the three reference genes included; *HBB*, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and actin beta (*ACTB)*. The probe features of selected genes and PCR conditions are shown in Table 2.10. Reference genes are used for normalization for sample differences in cDNA concentrations. Therefore, stable and constant reference genes are required for normalization. Three reference genes were selected which showed the same differentiated expression pattern in both blood and brain tissues. They were selected based on the published literature in which these genes were used in gene expression studies on blood and brain and by searching the Ref genes database (Hruz et al., 2011) (http://www.refgenes.org/rg/).

**Table 2.10: Real time PCR primers and conditions**

| Gene | Product Size (bp) | Annealing (˚C) | Gene ID | GenBank Accession | Catalogue Number |
|---|---|---|---|---|---|
| *CEBPA* | 88 | 60 | 1050 | NM_004364 | QT00203357 |
| *HBB* | 105 | 60 | 3043 | NM_000518 | QT00244489 |
| *GAPDH* | 95 | 60 | 2597 | NM_001289745 | QT00079247 |
| *ACTB* | 104 | 60 | 60 | NM_001101 | QT01680476 |

*CEBPA*, CCAAT/enhancer binding protein alpha; *HBB*, hemoglobin subunit beta; *GAPDH*, glyceraldehyde-3-phosphate dehydrogenase; *ACTB*, actin beta.

QuantiTect®Primer assays were reconstituted by adding 1.1ml 1X TE, pH 8.0 to each vial and vortexed. Primer assays were aliquoted into separate tubes and stored at -20˚C until required.

### 2.8.2 RNA isolation

All study participant RNA previously isolated (Section 2.5.2) and stored at -80˚C was used for qPCR analysis.

### 2.8.3 cDNA Synthesis

Isolated RNA from all study participants was converted to cDNA using the QuantiNova™ Reverse Transcription Kit (Protocol in section 2.5.5.2.1). In addition, non-reverse transcribed control samples (-RT) were prepared from previously isolated RNA using the QuantiNova™ Reverse Transcription Kit (Qiagen, Germany). This control serves to detect gDNA and other sources of contamination.

Briefly 1x gDNA wipeout buffer and RNase free water was added to 750 ng RNA and incubated at 45˚C for 2 minutes on the the 2720 Thermal Cycler (Applied Biosystems, CA, USA). Samples were then placed on ice. Reverse transcription mix was added at a final reaction volume of 18µl. Samples were incubated at 25˚C for 3 minutes for annealing, reverse transcription at 45˚C for 10 minutes followed by 85˚C for 5 minutes. All samples were placed on ice and subsequently stored at -80˚C if not required for immediate experiments.

The Nanodrop™ 2000 (Thermo Scientific, MA, USA) was used to determine the concentration and purity of all cDNA and –RT samples. The concentrations of cDNA and –RT samples were determined by measuring A260 absorbance for each sample. The purity was assessed using the A260/A230 ratio. All cDNA and –RT samples were diluted to 20ng/µl in 1x TE buffer and used for qPCR gene expression analysis.

### 2.8.4 Standard real time PCR

A total reaction master mix of 20µl for each sample (20 cases and 20 controls) was prepared which included: 1x QuantiTect Primer Assay (Qiagen, Hilden, Germany), 1x RNase free water, 1x FastStart Essential DNA Green Master Mix (Roche Diagnostics, Indianapolis, USA ) and 20 ng template cDNA (Table 2.11). The reagents were transferred into a 96 well plate using the epMotion®5070f (Eppendorf, Hamburg, Germany) and 50 µl filter epTIPS® (Eppendorf, Hamburg, Germany) were used to ensure a high degree of accuracy. All reactions were performed in triplicate. The plate was sealed and centrifuged at 20˚C for 1 minute at 1500 rpm on the Eppendorf 5810 R centrifuge *(*Eppendorf, Hamburg, Germany).

Centrifugation was performed to ensure no residual liquid remained on the sides of the plate wells and all contents were at the bottom of the wells.

The 96 well plates were run on the Lightcycler® 96 (Roche Diagnostics, Indianapolis, USA ) according to the thermo cycling profile in Table 2.12. A total of 45 cycles were used.

**Table 2.11: qPCR reaction master mix and set-up**

| Component | Concentration | Volume (µl) per reaction |
|---|---|---|
| FastStart Essential DNA Green Master | 1x | 10 |
| QuantiTect Primer Assay | 1x | 2 |
| Template cDNA | 20 ng | 1 |
| RNase free water | - | 7 |
| **Total Volume** | | **20µl** |

cDNA, complementary deoxyribonucleic acid; qPCR, quantitative real-time polymerase chain reaction

**Table 2.12: qPCR cycling conditions using the Lightcycler®96**

| PCR Step | Time (s) | Temperature (˚C) |
|---|---|---|
| **Preincubation** | 600 | 95 |
| **3-step activation (45 cycles)** | | |
| Denaturation | 10 | 95 |
| Annealing | 20 | 55 |
| Extension | 20 | 72 |
| **Melting** | | |
| | 10 | 95 |
| | 60 | 65 |
| | 1 | 95 |
| **Cooling** | 30 | 37 |

PCR, polymerase chain reaction

PCR conditions were optimized in a trial run and thereafter all samples were run, in triplicate according to the above mentioned protocol. An additional control sample, Human Foetal brain total RNA (Takara Bio Inc, CA, USA) was purchased and added for gene expression analysis.

### 2.8.5 Data Analysis

All data generated was analysed using the Lightcycler® 96 software version 1.1 (Roche Diagnostics, Indianapolis, USA) to generate quantification cycle (Cq) values. The Cq is defined as the cycle value at which the fluorescence of a sample represented by an amplification curve crosses the threshold. Data were normalized using a standardized method of qPCR analysis using geometric averaging in a spreadsheet in Microsoft Excel (Vandesompele et al., 2002).

For comparative purposes the relative gene expression approach was also utilized through Relative expression software tool (REST-2009 ©) version 2.0.13 (Qiagen, Hilden, Germany, http://rest.gene-quantification.info/). The tool is a mathematical model which includes several algorithms for comparsion of the target gene to multiple reference genes based on the real-time PCR efficiency and the mean crossing point deviation (delta Ct) between cases and control groups (Pfaffl et al., 2002).

# Chapter 3: Results

3.1 Screening for pathogenic mutations ..... 68
3.1.1 Sample Selection ..... 68
3.1.2 MLPA Analysis and Sanger sequencing ..... 68
3.1.2.1 MLPA Quality Control ..... 68
3.1.2.2 Detection of a heterozygous exon deletion in one patient ..... 69
3.2 RNA Isolation and Sequencing ..... 73
3.2.1 Determination of RNA concentration and purity by CAF ..... 73
3.2.2 In-house Quality Control to Detect gDNA Contamination ..... 76
3.2.3 External Quality Control by Nxt-Dx ..... 77
3.2.3.1 RNA QC to determine cDNA amount for input ..... 77
3.2.3.2 cDNA library preparation QC ..... 79
3.3 RNA-Seq Analysis Pipeline ..... 82
3.3.1 Quality Control- FastQC ..... 83
3.3.2 Alignment of Reads ..... 84
3.3.3 Differential Gene Expression ..... 85
3.3.4 Hierarchical Clustering ..... 90
3.4 Enrichment Analysis ..... 92
3.4.1 Gene Ontology ..... 92
3.4.2 Functional Annotation Clustering ..... 92
3.4.3 Ingenuity Pathway Analysis (IPA) ..... 93
3.5 Gene Prioritization ..... 98
3.6 Verification of *CEPBA* gene ..... 102

## 3.1 Screening for pathogenic mutations

### 3.1.1 Sample Selection

All PD patient and control samples met inclusion criteria as previously stated in the Methods section 2.2. Briefly, all patients had to have a disease duration of 4 or more years and the controls did not have PD or any other neurological disorder and were age, and ethnically matched to the patients. However, the cases and controls were not sex-matched, 40% of the PD patients were female whereas 100% of the controls were female. The mean ages at sampling of the PD patients and controls were 59.8 ± 9.9 and 57.2 ± 4.4 years, respectively (p-value of 0.076 and odds ratio of 1.11). The mean average age of onset of PD was 51.8 ± 9.8 years. All study participants self-identified as being of the mixed ancestry ethnic group. All 20 PD patients were screened for pathogenic mutations using the MLPA technique.

### 3.1.2 MLPA Analysis and Sanger sequencing

MLPA analysis can detect single base pair changes as well as exonic rearragements in candidate genes. The MLPA kits P051 and P052 utilized in the present study contained probes specific to PD which allowed for detection of exonic deletions or duplications in *SNCA*, *PARK2* (*Parkin*), Ubiquitin C-Terminal Hydrolase L1 *(UCHL1),* PTEN-induced putative kinase 1 *(PINK1), PARK7, LRRK2, GCH1* , *ATP13A2* and two point mutations ( *SNCA* A30P and *LRRK2* G2019S).

#### 3.1.2.1 MLPA Quality Control

SALSA MLPA probe mixes contain nine control fragments including X & Y chromosome probes. For all patients, the genders from both the P051 and P052 kits correlated with the demographic information provided in Table 2.1.

**Quantity control fragments (Q-fragments)**

Q-fragments which include 64, 70, 76 and 80 nt fragments are sensitive to the DNA concentration and ligation efficiency. These fragments are not hybridized, ligated or amplified during the PCR run. Therefore the higher amount of template DNA used, the lower the Q-fragments, and vice versa. Too high concentrations of DNA caused signal sloping as the samples could contain contaminants which inhibit polymerase. Therefore, we optimized the experiment and found that 100 ng of DNA per sample was ideal as this amount reduced contaminants and generated good MLPA peak profiles that could be used for MLPA copy number detection.

**Denaturation control fragments (D-fragments)**

D-fragments (88 and 96 nt fragments) provide warning about denaturation problems. Initially, the 88nt D-fragment was low on the electropherogram in some samples which indicated problems with hybridization, and this was due to the fact that initially reaction volumes had been halved to save on costs. Using full reaction volumes and the recommended protocol (provided in the Materials and methods section 2.4.3) resolved these problems.

**General Quality Control**

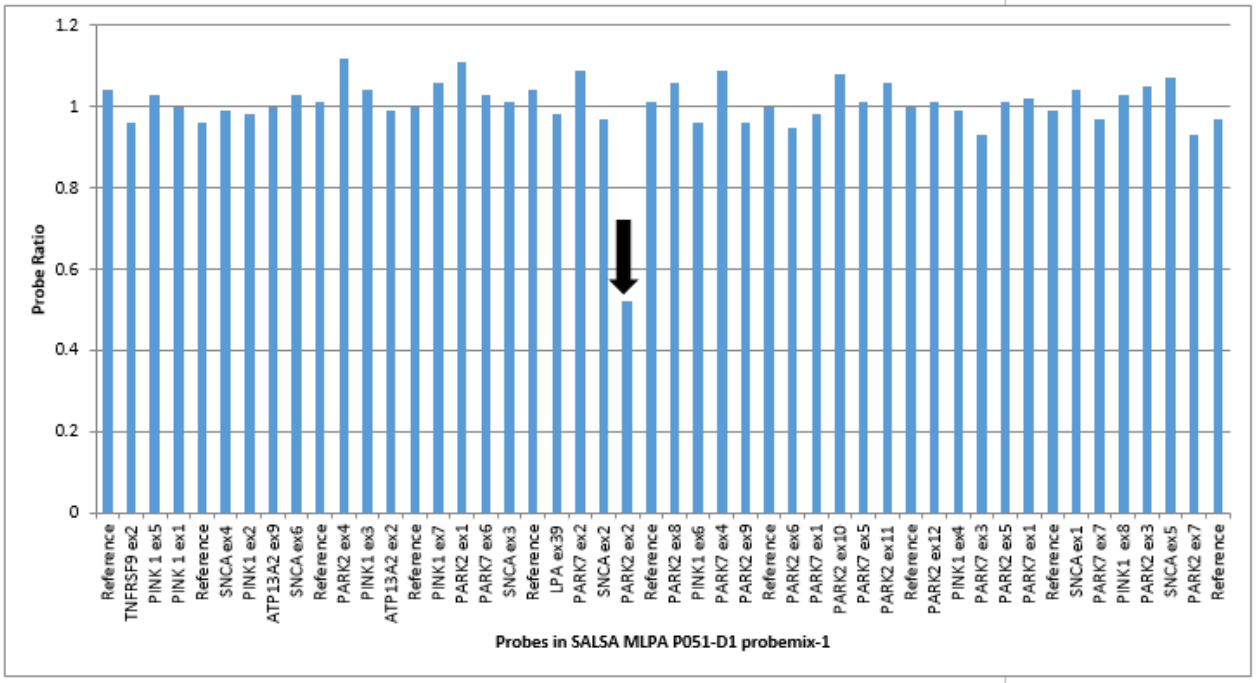
Three reference DNA samples using the same DNA extraction method were also included in each run to ensure good reproducibility of each probe within each MLPA run. These reference samples were selected from the 20 controls. Additionally, a no DNA control was included in each MLPA run to ensure there was no contamination present in the run.

Samples which did not display acceptable internal control fragments after troubleshooting were repeated in a second MLPA run. Upon rerun, all patient samples displayed appropriate Q and D-fragment profiles.

### 3.1.2.2 Detection of a heterozygous exon deletion in one patient

MLPA peak profiles were generated for all patient samples and are represented as electrophoregrams (Appendix XIV, Figure 1). However, Copy number variations (CNVs) cannot be determined by using the visual interpretation of an electrophoregram, therefore statistical analysis using reference samples and regression lines was required to determine CNVs and Coffalyser.net software (MRC, Holland) was used for this purpose. A heterozygous exon 2 deletion in the *PARK2* gene was detected in one male patient sample, SR023. This finding was observed with both MLPA kits P051 (Figure 3.1A) and P052 (Figure 3.1B), as probes for *PARK2* are present in both kits. This finding was verified using qPCR (Figure 3.2).

**A**




**B**

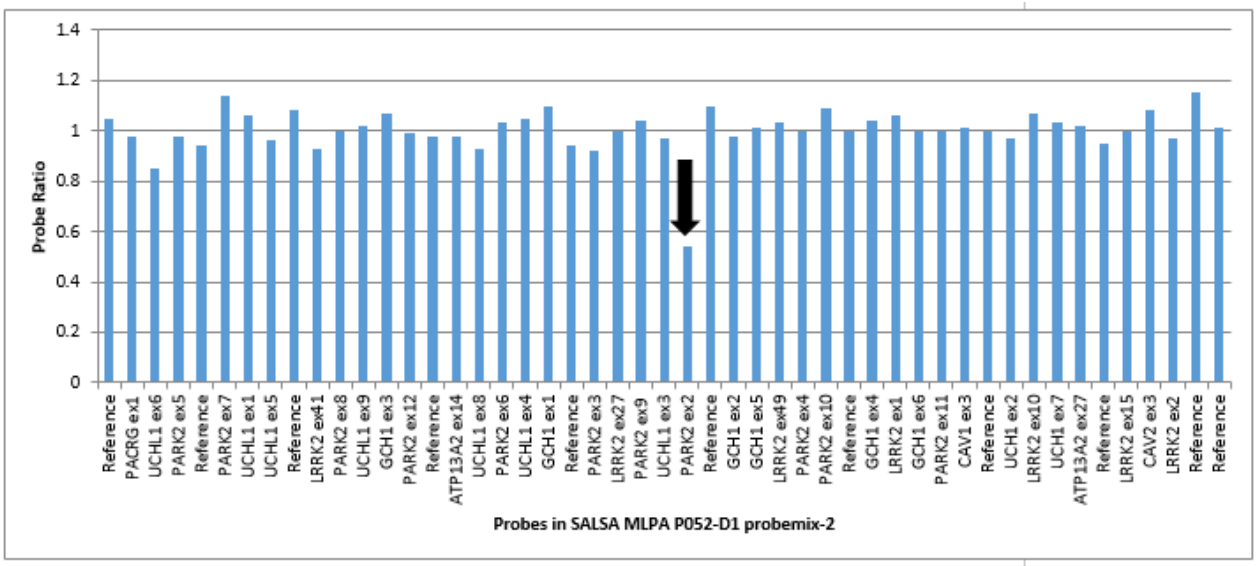



**Figure 3.1: MLPA analysis results from a male PD patient SR023 with a heterozygous deletion of PARK2 exon 2 (ratio<0.6).** Bar graphs representing the normalized ratios for each exon using the MLPA Mix 1 P051 (A) and Mix 2 P052 (B). Black arrows indicate the heterozygous *PARK2* exon 2 deletion.

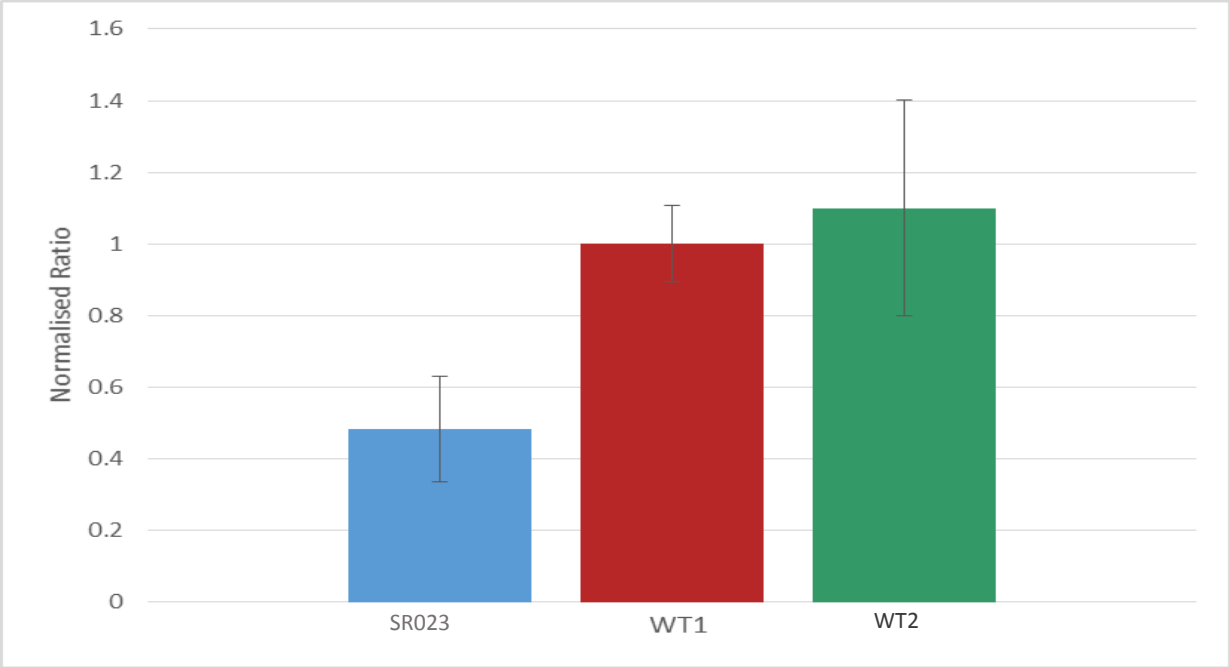
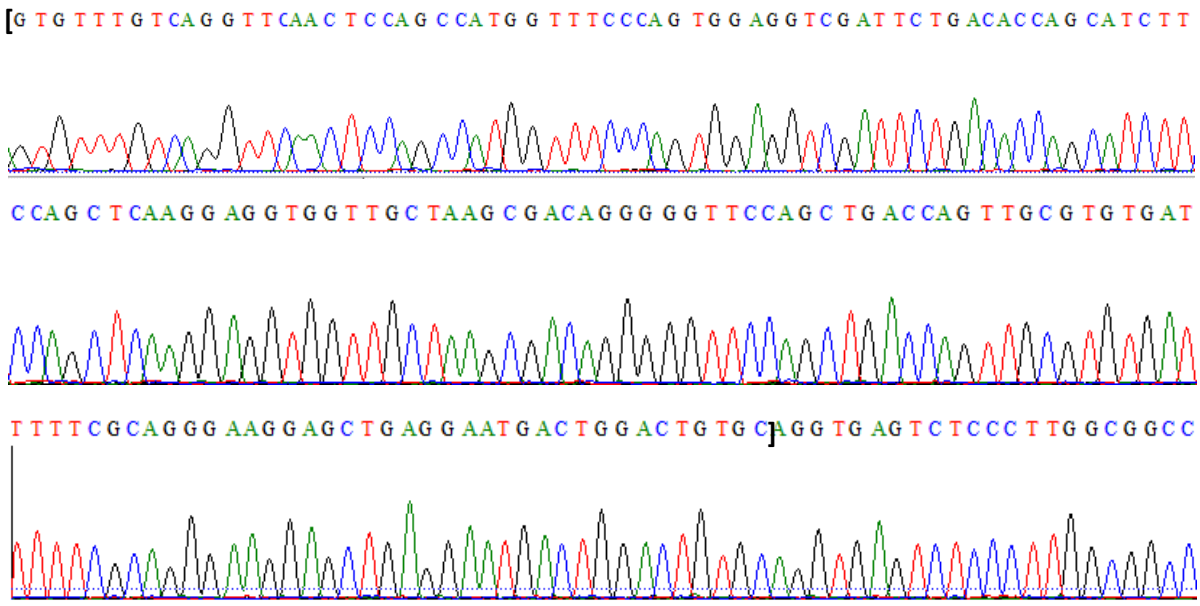



**Figure 3.2: Quantitative real-time PCR verified a *PARK2* exon 2 heterozygous deletion in patient SR023.** *PARK2* exon 2 was present in only one allele of SR023 compared to controls (WT1 and WT2). SR023 and wild-type samples were normalized to reference gene, haemoglobin subunit beta (*HBB)*. WT, wild type

Next, Sanger sequencing of all exons of *PARK2* was performed in patient SR023 to detect a possible second pathogenic mutation in this gene. Mutations in *PARK2* are known to cause an autosomal recessive form of PD and therefore a second pathogenic mutation needs to be identified to confirm that *PARK2* is the cause of PD in this patient. However, sequencing of all 12 exons of *PARK2* revealed no other pathogenic mutations in this patient. The sequencing chromatogram for *PARK2* exon 2 is shown in Figure 3.3 showing clean sequencing of good quality and that no variants were detected in this exon.

**A**




**B**

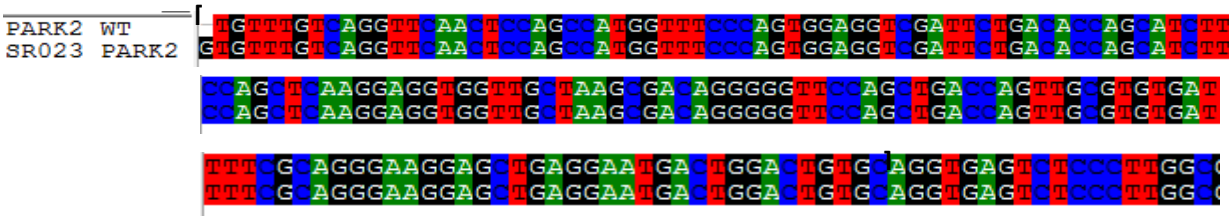



**Figure 3.3: Sanger sequencing of *PARK2* exon 2**. A) Chromatogram of *PARK2* exon 2 for SR023 illustrating no mutations were identified. Start and end positions of *PARK2* exon 2 are indicated by square brackets B) Sequence alignment of *PARK2* exon 2 from patient SR023 to the *PARK2* reference sequence. WT, wild-type.

Since no pathogenic mutations were identified in any of the known PD-causing genes using MLPA, this means that all 20 patients potentially have idiopathic PD, and therefore could be analyzed as one group for the RNA-Seq analysis.

## 3.2 RNA Isolation and Sequencing

### 3.2.1 Determination of RNA concentration and purity by CAF

RNA concentration and integrity of all patient and control samples was determined using the Agilent 2100 Bioanalyser at CAF (Table 3.1). All of the RNA extracted produced RIN values of greater than 7 which indicated that the RNA was intact, of good quality and suitable for downstream experimental purposes. The Agilent Bioanalyser software generates an electropherogram and virtual gel image of the samples to detect the extent of RNA degradation and provides an assessment of RNA quality. Degraded RNA is expected to produce a decreased 18S and 28S ribosomal RNA (rRNA) ratio, an increased baseline signal, and a shift to shorter fragment lengths on the virtual gel. A representative virtual gel is shown in Figure 3.4A, where the green line denotes the 25bp ladder and the two darker bands represent the 18S and 28S rRNA peaks illustrating RNA of good quality with minimal degradation. A representative electropherogram is also included showing the 18S and 28S peaks for a sample with a RIN value of 9.4 (Figure 3.4 B). Sample SR046 had the lowest RIN value (RIN 7.3) compared to other samples. This could be due to the fact that during RNA extraction, the blood in the tube had been very viscous and this could have affected the quality of the RNA. Overall no significant difference was observed between patients and controls with the average RIN value of 9.3 for controls and 9.0 for patients.

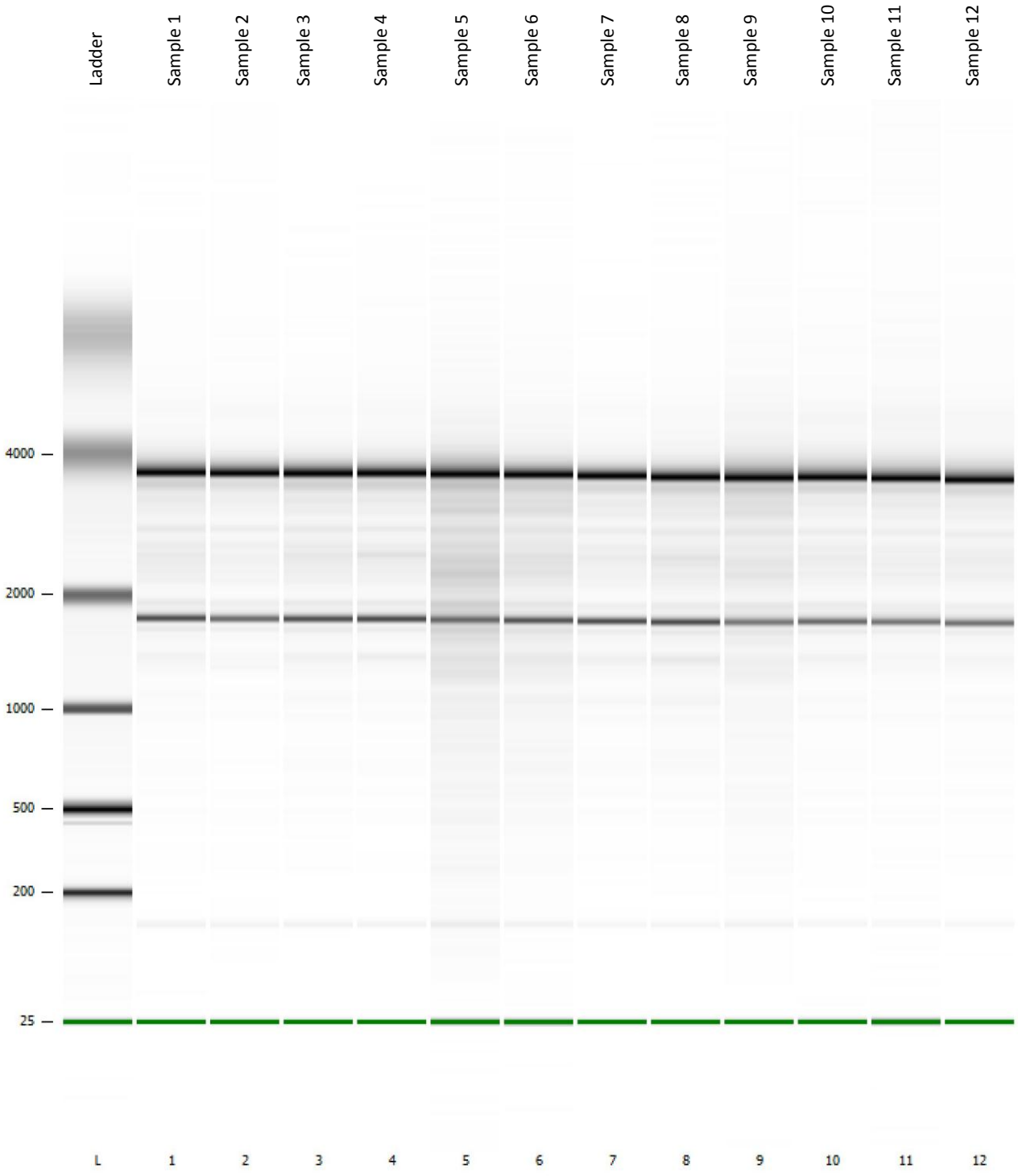
**Table 3.1: RNA concentration and purity for patient and control blood samples collected in PAXgene tubes using the Agilent 2100 Bioanalyser.**

| Sample ID | Gender | Clinical Status | Disease Duration (yrs) | RNA area under the curve | RNA Concentration (ng/µl) | RNA ratio [28s/18s] | RIN value |
|---|---|---|---|---|---|---|---|
| SR013 | F | HC | N/A | 98.8 | 53 | 2.3 | 9.6 |
| SR098 | F | HC | N/A | 56.6 | 30 | 2.3 | 7.8 |
| SR050 | F | HC | N/A | 137 | 73 | 2.7 | 9.7 |
| SR079 | F | HC | N/A | 136.7 | 73 | 2.6 | 9.5 |
| SR055 | F | HC | N/A | 223.7 | 119 | 2.3 | 9.6 |
| SR103 | F | HC | N/A | 171.3 | 91 | 2.6 | 9.6 |
| SR006 | F | HC | N/A | 69.1 | 37 | 3.0 | 9.3 |
| SR089 | F | HC | N/A | 122.2 | 65 | 2.7 | 9.3 |
| SR082 | F | HC | N/A | 87.6 | 47 | 2.7 | 9.6 |
| SR109 | F | HC | N/A | 139.9 | 96 | 2.5 | 9.6 |
| SR119 | F | HC | N/A | 168.1 | 115 | 2.5 | 9.9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SR099 | F | HC | N/A | 361.2 | 248 | 2.3 | 9.4 |
| SR111 | F | HC | N/A | 268.5 | 184 | 2.4 | 9.3 |
| SR080 | F | HC | N/A | 105.9 | 73 | 2.9 | 9.3 |
| SR093 | F | HC | N/A | 260.2 | 179 | 2.5 | 9.3 |
| SR088 | F | HC | N/A | 323.5 | 222 | 2.4 | 9.4 |
| SR092 | F | HC | N/A | 353.6 | 243 | 1.1 | 8.0 |
| SR054 | F | HC | N/A | 290.9 | 200 | 2.9 | 9.4 |
| SR086 | F | HC | N/A | 478.4 | 328 | 2.2 | 9.0 |
| SR027 | F | HC | N/A | 545.1 | 299 | 2.5 | 9.4 |
| SR053 | M | PD | 6 | 192.4 | 132 | 2.7 | 8.9 |
| SR033 | M | PD | 4 | 272.8 | 187 | 2.7 | 9.5 |
| SR002 | M | PD | 5 | 256.4 | 137 | 2.2 | 9.5 |
| SR005 | F | PD | 8 | 162.6 | 87 | 2.3 | 9.8 |
| SR121 | M | PD | 15 | 269 | 143 | 2.5 | 9.4 |
| SR011 | F | PD | 8 | 434.8 | 232 | 2.0 | 9.2 |
| SR046 | F | PD | 10 | 165.4 | 88 | 2.6 | 7.3 |
| SR023 | M | PD | 4 | 121.2 | 65 | 2.5 | 8.4 |
| SR009 | F | PD | 12 | 179.8 | 96 | 1.7 | 9.1 |
| SR056 | M | PD | 21 | 212.5 | 113 | 2.5 | 8.8 |
| SR070 | M | PD | 7 | 248.2 | 132 | 2.6 | 8.8 |
| SR041 | F | PD | 7 | 171.8 | 92 | 3.0 | 9.4 |
| SR039 | F | PD | 6 | 68.7 | 37 | 2.5 | 9.6 |
| SR014 | M | PD | 9 | 173.9 | 93 | 2.4 | 9.4 |
| SR042 | M | PD | 4 | 401.5 | 220 | 1.1 | 8.7 |
| SR067 | F | PD | 7 | 93.2 | 51 | 2.9 | 7.9 |
| SR012 | M | PD | 5 | 186 | 102 | 2.8 | 9.4 |
| SR007 | M | PD | 4 | 135.8 | 74 | 2.4 | 9.5 |
| SR010 | F | PD | 13 | 43.9 | 24 | 2.8 | 10 |
| SR003 | M | PD | 6 | 207.8 | 114 | 2.5 | 9.3 |

F: Female; M: Male; HC: healthy control; PD: Parkinson's disease; N/A: not applicable; RIN: RNA integrity number; yrs, years

**A**



**B**

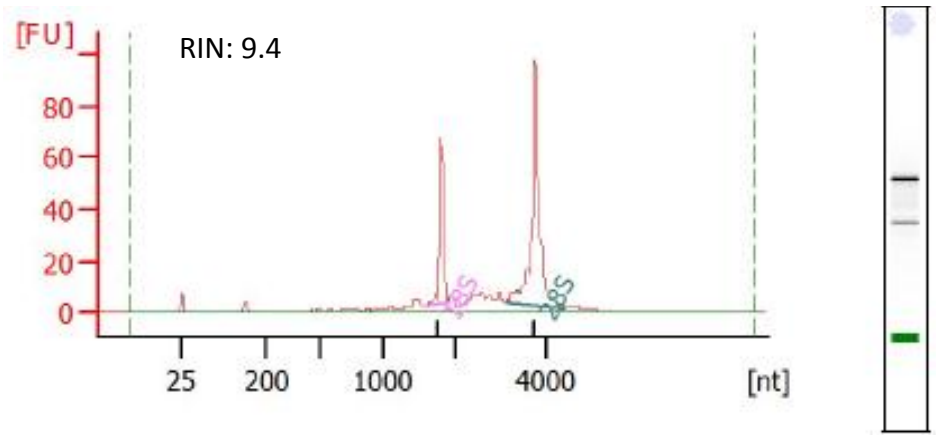



**Figure 3.4 RNA quality and concentration as assessed by the Agilent 2100 Bioanalyser.** A) A representative virtual RNA gel showing the results of RNA integrity of the patient and control samples. The lowest bands denoted by the green line represent the 25bp ladder whereas the two darker bands represent the 28S and 18S rRNA peaks. B) Representative electropherogram showing the 18S and 28S RNA peaks and illustrating good quality RNA. FU, fluorescence units, nt, nucleotides

### 3.2.2 In-house Quality Control to Detect gDNA Contamination

As an additional in-house quality control test, RNA extracted from a control blood sample (SR013) was reverse transcribed to cDNA. PCR was then performed using the HBB primer set which produces different sized PCR fragments depending on whether gDNA or cDNA is present (as described in Figure 2.3 in the Materials and methods section). As expected, PCR products of 377bp for the gDNA controls (positive controls) and 247bp for SR013 control sample were obtained (Figure 3.5). These results confirm successful RNA extraction, with negligible gDNA contamination, and successful cDNA conversion for the RNA sample.

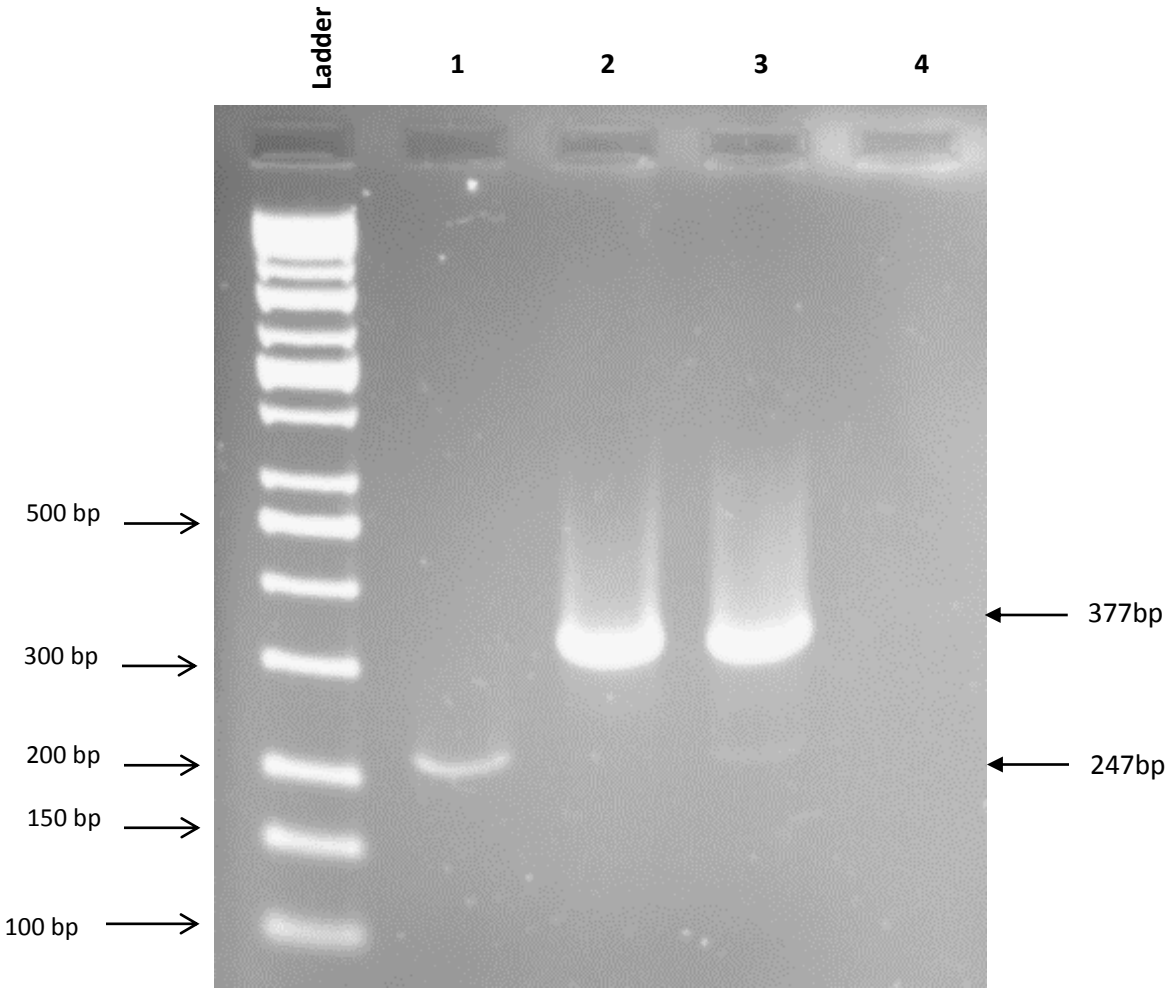



**Figure 3.5: Amplification of cDNA and gDNA templates using the *HBB* primer set.** The Kapa universal ladder was used. Lane 1 is a control sample (SR013) which produced a PCR product of 247bp. Lanes 2 and 3 are gDNA templates which produced PCR products of 377bp, and therefore served as positive controls in this experiment. Lane 4 is a negative control containing distilled water which does not amplify and therefore no band is visible.

### 3.2.3 External Quality Control by Nxt-Dx

#### 3.2.3.1 RNA Quality control to determine cDNA amount for input

RNA quality and purity was determined using the Agilent 2100 Bioanalyser by NXT-Dx (Gent, Belgium) before cDNA library preparation was performed (Table 3.2). RIN values from CAF and Nxt-DX were equivalent in 6 samples namely; SR099, SR080, SR092, SR027, SR067 and SR012. Only one sample (SR039) indicated variation between RIN values of greater than 1. Therefore, overall the RIN values from CAF and Nxt-Dx were well correlated. Additionally, the Agilent Eukaryote Total RNA Pico Assay was used by NXT-Dx to determine RNA concentrations of PD patient and control samples (Table 3.2).

**Table 3.2: RNA concentration, purity using the Agilent 2100 Bioanalyser. Comparison of the RIN values between internal (CAF) and external (Nxt-Dx) quality control**

| Sample ID | RNA concentration (ng/µl) (Nxt-Dx) | RIN value (Nxt-Dx) | RIN value (CAF) |
|---|---|---|---|
| SR013 | 86.0 | 9.1 | 9.6 |
| SR098 | 33.3 | 7.7 | 7.8 |
| SR050 | 54.1 | 8.9 | 9.7 |
| SR079 | 103.5 | 9.3 | 9.5 |
| SR055 | 114.8 | 9.4 | 9.6 |
| SR103 | 76.7 | 9.2 | 9.6 |
| SR006 | 36.8 | 9.1 | 9.3 |
| SR089 | 69.9 | 8.9 | 9.3 |
| SR082 | 48.0 | 9.3 | 9.6 |
| SR109 | 75.9 | 9.3 | 9.6 |
| SR119 | 109.6 | 9.5 | 9.9 |
| SR099 | 246.0 | 9.4 | 9.4 |
| SR111 | 170.6 | 9.0 | 9.3 |
| SR080 | 67.3 | 9.3 | 9.3 |
| SR093 | 157.8 | 9.0 | 9.3 |
| SR088 | 181.2 | 8.9 | 9.4 |
| SR092 | 197.5 | 8.0 | 8.0 |
| SR054 | 156.1 | 9.0 | 9.4 |
| SR086 | 269.1 | 9.3 | 9.0 |
| SR027 | 251.1 | 9.4 | 9.4 |
| SR053 | 103.0 | 9.5 | 8.9 |
| SR033 | 112.2 | 9.7 | 9.5 |
| SR002 | 125.5 | 9.6 | 9.5 |
| SR005 | 97.3 | 9.7 | 9.8 |

| | | | |
|---|---|---|---|
| SR121 | 165.0 | 9.0 | 9.4 |
| SR011 | 239.2 | 9.5 | 9.2 |
| SR046 | 80.5 | 8.2 | 7.3 |
| SR023 | 69.2 | 8.7 | 8.4 |
| SR009 | 88.3 | 8.2 | 9.1 |
| SR056 | 95.0 | 9.3 | 8.8 |
| SR070 | 106.7 | 9.0 | 8.8 |
| SR041 | 72.8 | 9.5 | 9.4 |
| SR039 | 37.2 | 8.2 | 9.6 |
| SR014 | 69.5 | 9.7 | 9.4 |
| SR042 | 246.7 | 8.3 | 8.7 |
| SR067 | 69.0 | 7.9 | 7.9 |
| SR012 | 86.3 | 9.4 | 9.4 |
| SR007 | 46.5 | 9.5 | 9.5 |
| SR010 | 27.3 | 9.2 | 10 |
| SR003 | 131.1 | 9.8 | 9.3 |

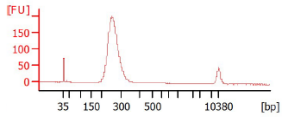
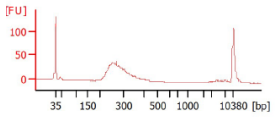
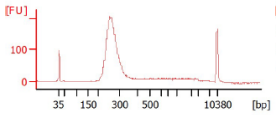
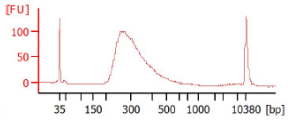
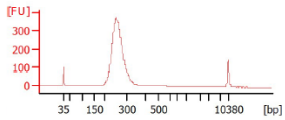
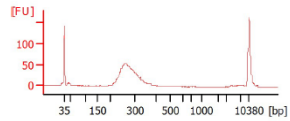
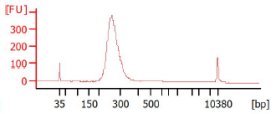
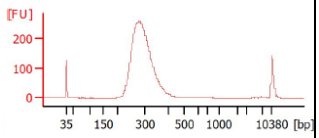
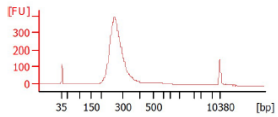
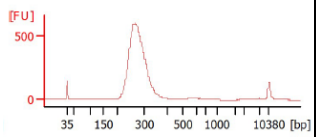
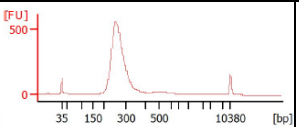
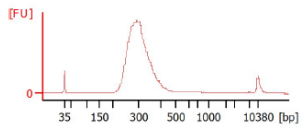
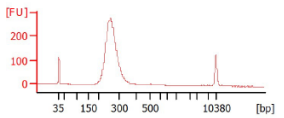
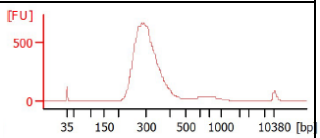
CAF, central analytical facility; RIN, RNA intergrity number

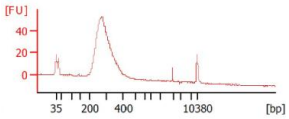
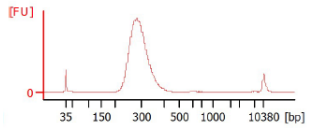
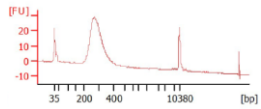
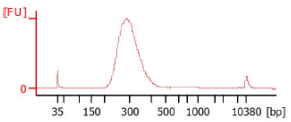
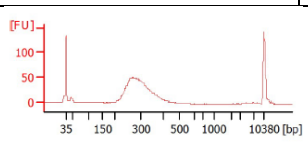
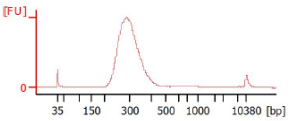
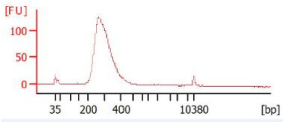
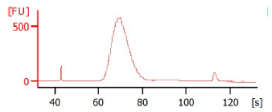
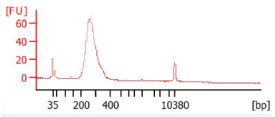
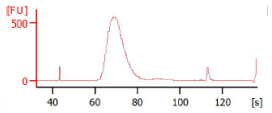
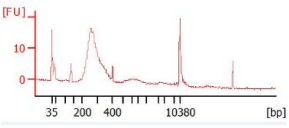
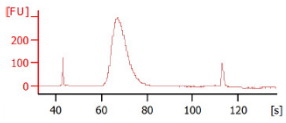
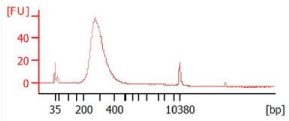
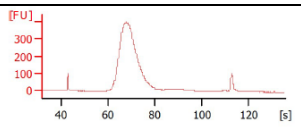
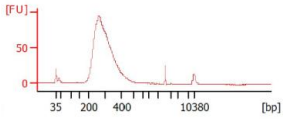
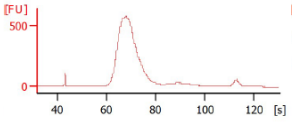
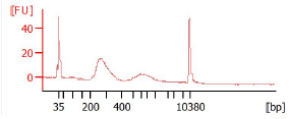
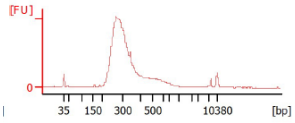
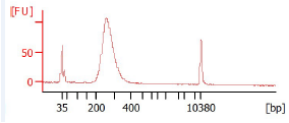
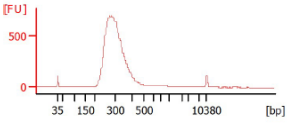
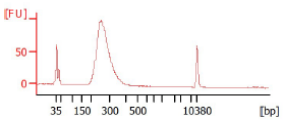
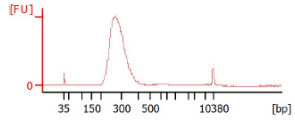
#### 3.2.3.2 cDNA library preparation Quality Control

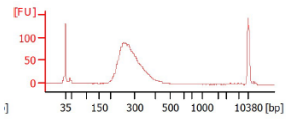
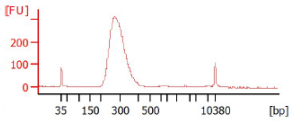
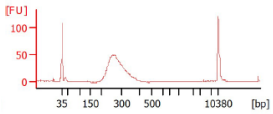
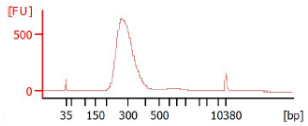
A total of 550 ng RNA (Appendix XII, Table 6) was required for all RNA samples before conversion to cDNA. After RNA fragmentation, cDNA libraries were prepared for each sample using TruSeq Stranded Total RNA Library Prep Kit and quality control on these was performed on the Agilent 2100 HS DNA Chip (Table 3.3). The concentration was determined with smear analysis on the Agilent 2100. All electropherograms showed a narrow distribution with a peak size at approximately 300 bp which is the optimal input size distribution for the TruSeq Stranded Total RNA Library Prep Kit. This is an important step in the RNA-Seq procedure because if this quality control (QC) fails then the reliability of the RNA sequencing reads would be questionable. The cDNA electropherogram should have three visible peaks. Two standard peaks namely the lower marker and the upper marker and the sample peak at 300 bp. The lower marker is visible at 35 bp and is used for alignment and the upper marker used for quantification and visible at 1500 bp. The x-axis can be changed from time (in seconds) to size (in bp) or vice versa which is visible in the cDNA electropherograms however this does not change the peak profile. Whereas

the y-axis shows the signal intensity. All cDNA samples passed QC and could therefore be sequenced on the Illumina HiSeq 4000 (Illumina Inc, San Diego, USA).

**Table 3.3: cDNA concentration and quality during cDNA library preparation.**

| Sample ID of Controls | cDNA library Concentration (pg/µl) | Electropherogram of cDNA library | Sample ID of Cases | cDNA library Concentration (pg/µl) | Electropherogram of cDNA library |
|---|---|---|---|---|---|
| SR013 | 1649.73 |  | SR053 | 478.50 |  |
| SR098 | 1851.42 |  | SR033 | 1503.39 |  |
| SR050 | 3596.56 |  | SR002 | 448.18 |  |
| SR079 | 3532.43 |  | SR005 | 2597.56 |  |
| SR055 | 3777.77 |  | SR121 | 5114.00 |  |
| SR103 | 4731.09 |  | SR011 | 5041.40 |  |
| SR006 | 2412.50 |  | SR046 | 8131.57 |  |

| | | | | | |
|---|---|---|---|---|---|
| SR089 | 2021.42 |  | SR023 | 4165.41 |  |
| SR082 | 181.66 |  | SR009 | 3882.91 |  |
| SR109 | 692.33 |  | SR056 | 7045.33 |  |
| SR119 | 145.91 |  | SR070 | 6932.14 |  |
| SR099 | 67.70 |  | SR041 | 6129.04 |  |
| SR111 | 1489.81 |  | SR039 | 3175.91 |  |
| SR080 | 487.04 |  | SR014 | 5648.92 |  |
| SR093 | 1851.42 |  | SR042 | 13039.67 |  |
| SR088 | 2412.50 |  | SR067 | 8886.85 |  |
| SR092 | 1592.89 |  | SR012 | 11578.88 |  |
| SR054 | 1662.96 |  | SR007 | 6608.54 |  |

| | | | | | |
|---|---|---|---|---|---|
| SR086 | 1212.54 |  | SR010 | 4290.89 |  |
| SR027 | 678.24 |  | SR003 | 7077.61 |  |

cDNA, complementary deoxyribonucleic acid

## 3.3 RNA-Seq Analysis Pipeline

RNA-Seq data analysis was performed using Partek® Flow® Software build 4.0.15 (Partek Incorporated, St Louis, MO, USA). Unaligned reads underwent quality control using FastQC, reads were then aligned using STAR, transcriptome quantification was performed and a candidate gene list of differentially expressed genes was generated. Figure 3.6 shows an overview of the pipeline and the bioinformatics tools employed in the present study.

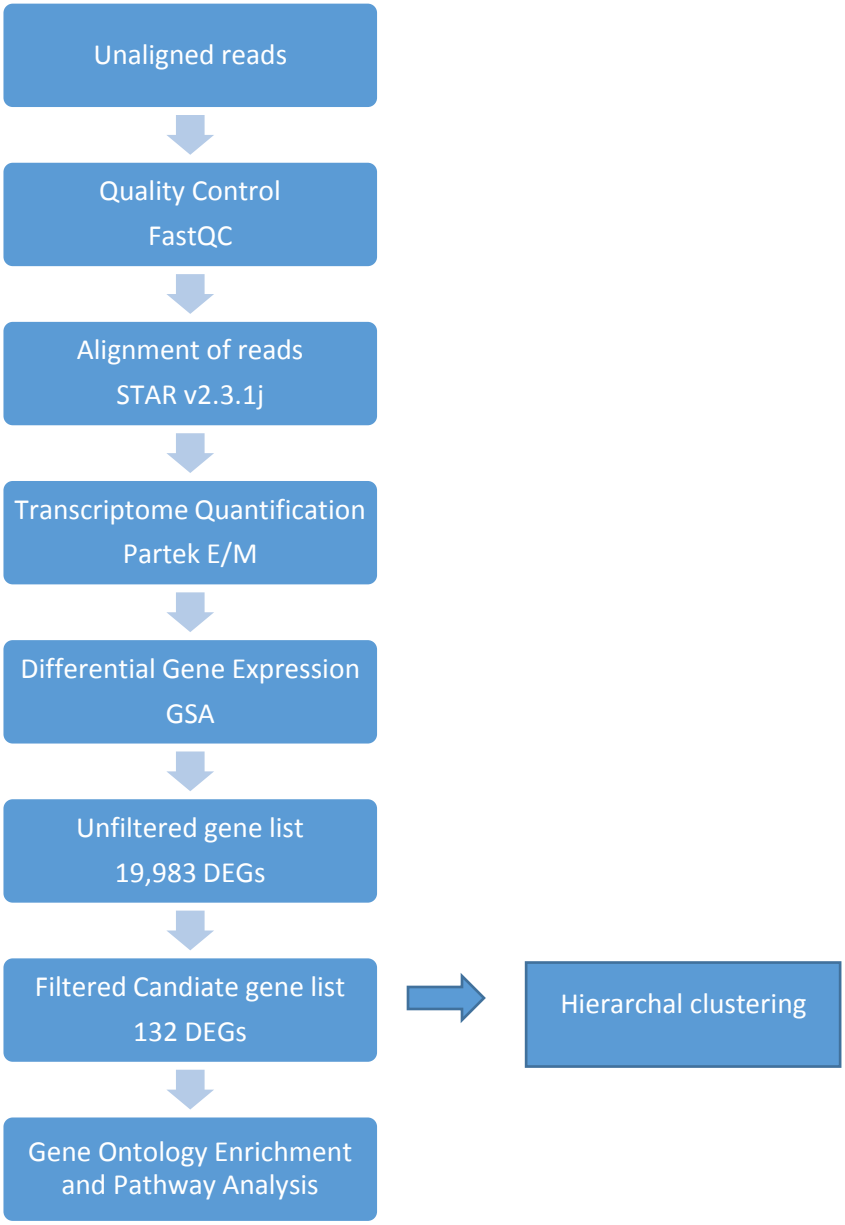



**Figure 3.6: Flowchart of the RNA-Seq data analysis pipeline.** QC was performed on the raw sequencing reads using FastQC, filtered reads were mapped to a reference file hg19 GENCODE using STAR, Partek E/M algorithm was used for transcriptome quantification and GSA was used for differential gene expression analysis which generated an unfiltered gene list, which was then filtered down and hierarchal clustering was performed as well as enrichment analysis. DEGs, differentially expressed genes.

### 3.3.1 Quality Control- FastQC

The FASTQC quality control tool was used for an initial quality assessment. The quality scores across the bases for each sample were generated and a representative sample is provided in Appendix XV, Table 8. The report generated showed the range of quality values across all bases at each position. The x-axis

gives the read position and the y-axis represents the quality scores. The higher the score on the y-axis the better the base call. All samples had good quality scores indicated by the green region which divided the y-axis. The average quality of all reads was above 30. Quality trimming removed 8 bp and resulted in reads of approximately 42 bp.

### 3.3.2 Alignment of Reads

Reads were aligned using STAR v2.3.1j. Table 3.4 shows the number of reads and mapping for each sample. The average read length for all samples ranged from 29 million (M) to 55 M reads. The. average quality scores were above 39 for all samples. Coverage ranged from 7% to 36% and the average percentage alignment was 95.3% for samples.

**Table 3.4: Total number of reads and read mapping for patient and control samples.**

| Sample name | Total reads | Total alignments | Aligned | Coverage | Average coverage depth | Average quality |
|---|---|---|---|---|---|---|
| SR013 | 32,685,114 | 83,428,740 | 94.02% | 13.61% | 9.83 | 39.40 |
| SR098 | 29,816,146 | 78,463,454 | 93.71% | 13.33% | 9.44 | 39.43 |
| SR050 | 42,024,983 | 109,353,282 | 94.50% | 14.97% | 11.72 | 39.48 |
| SR079 | 38,986,534 | 102,839,908 | 95.40% | 14.58% | 11.32 | 39.46 |
| SR055 | 36,769,630 | 93,734,242 | 93.13% | 14.31% | 10.50 | 39.22 |
| SR103 | 49,188,860 | 124,148,422 | 94.93% | 16.82% | 11.83 | 39.42 |
| SR006 | 42,276,367 | 96,725,860 | 93.75% | 16.40% | 9.46 | 39.44 |
| SR089 | 37,277,420 | 85,118,488 | 94.91% | 7.75% | 17.63 | 39.36 |
| SR109 | 43,474,282 | 103,189,816 | 95.68% | 8.36% | 19.81 | 39.38 |
| SR119 | 31,427,071 | 72,306,802 | 95.95% | 10.59% | 10.96 | 39.41 |
| SR099 | 37,643,726 | 94,493,586 | 95.93% | 8.88% | 17.08 | 39.46 |
| SR111 | 36,200,000 | 91,814,732 | 96.90% | 7.71% | 19.12 | 39.66 |
| SR080 | 44,736,939 | 112,263,166 | 96.61% | 9.54% | 18.89 | 39.53 |
| SR093 | 46,184,033 | 109,253,856 | 96.56% | 12.00% | 14.62 | 39.53 |
| SR088 | 41,396,232 | 101,273,694 | 96.72% | 5.04% | 32.19 | 39.55 |
| SR092 | 35,166,755 | 86,341,574 | 96.82% | 8.80% | 15.73 | 39.45 |
| SR054 | 40,606,773 | 96,230,784 | 96.73% | 8.92% | 17.30 | 39.60 |
| SR086 | 50,883,299 | 111,845,230 | 95.00% | 10.03% | 17.88 | 39.53 |
| SR027 | 35,074,377 | 77,870,220 | 95.81% | 9.29% | 13.45 | 39.58 |
| SR053 | 39,117,148 | 94,240,442 | 95.64% | 9.76% | 15.49 | 39.28 |
| SR033 | 34,440,746 | 80,565,174 | 94.41% | 11.05% | 11.70 | 39.32 |
| SR002 | 40,971,853 | 108,833,218 | 96.45% | 7.15% | 24.42 | 39.49 |
| SR005 | 42,770,701 | 99,818,832 | 96.19% | 17.15% | 9.35 | 39.49 |
| SR121 | 38,436,198 | 107,948,402 | 95.67% | 12.81% | 13.50 | 39.43 |
| SR011 | 39,054,532 | 113,855,220 | 95.30% | 11.20% | 16.26 | 39.12 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SR046 | 38,901,245 | 99,128,940 | 94.83% | 16.76% | 9.48 | 39.28 |
| SR023 | 45,273,148 | 114,061,586 | 95.84% | 16.14% | 11.32 | 39.31 |
| SR009 | 39,009,622 | 98,075,330 | 96.08% | 15.45% | 10.18 | 39.20 |
| SR056 | 35,384,607 | 93,952,010 | 95.76% | 14.37% | 10.47 | 39.22 |
| SR070 | 41,167,469 | 98,478,496 | 95.29% | 15.90% | 9.93 | 39.18 |
| SR041 | 33,082,683 | 78,405,442 | 95.68% | 14.31% | 8.78 | 39.31 |
| SR039 | 53,700,068 | 124,389,994 | 93.96% | 17.00% | 11.74 | 39.36 |
| SR014 | 52,783,964 | 133,880,932 | 96.07% | 17.13% | 12.53 | 39.36 |
| SR042 | 45,306,153 | 129,436,540 | 96.03% | 14.55% | 14.26 | 39.35 |
| SR067 | 45,615,167 | 111,308,590 | 90.71% | 36.87% | 4.82 | 39.36 |
| SR012 | 52,222,517 | 129,279,670 | 95.13% | 17.48% | 11.87 | 39.39 |
| SR007 | 43,902,418 | 115,845,278 | 93.76% | 13.49% | 13.77 | 39.49 |
| SR010 | 54,644,317 | 135,711,398 | 94.14% | 15.53% | 14.02 | 39.45 |
| SR003 | 50,536,139 | 123,304,180 | 96.54% | 17.29% | 11.43 | 39.42 |

### 3.3.3 Differential Gene Expression

Differential gene expression analysis was performed using the gene specific analysis algorithm (GSA) and this resulted in 19,983 unfiltered differentially expressed genes in the dataset.

Filtering parameters were set to a fold-change of ≤-2 or ≥+2, p-value of ≤ 0.01 and FDR of ≤ 0.05 to reduce the chances of false positives. These parameters resulted in a filtered candidate gene list of 132 differentially expressed genes (Table 3.5). The total reads ranged from 125 to 16,500.

**Table 3.5: Candidate gene list of differentially expressed genes in PD patients compared to controls using specified filtering parameters.**

| Gene symbol | Total reads | P-value (Cases vs. Controls) | FDR step up (Cases vs. Controls) | Fold change (Cases vs. Controls) |
|---|---|---|---|---|
| *LINC00278* | 1320 | 0.000000192 | 0.000167 | 20300 |
| *RPS4Y1* | 14000 | 0.00000000325 | 0.0000072 | 12200 |
| *EIF1AY* | 8000 | 0.0000000482 | 0.0000642 | 6290 |
| *TXLNG2P* | 16500 | 0.000000117 | 0.000115 | 3350 |
| *USP9Y* | 13600 | 0.0000000769 | 0.0000904 | 3330 |
| *TMSB4Y* | 443 | 0.000000988 | 0.000569 | 3280 |
| *RP11-576C2.1* | 125 | 0.0000243 | 0.00399 | 2700 |

| | | | | |
|---|---|---|---|---|
| *BCORP1* | 1460 | 0.00000139 | 0.000657 | 2550 |
| *TTTY10* | 295 | 0.00000481 | 0.00131 | 2070 |
| *RP11-424G14.1* | 439 | 0.0000047 | 0.00131 | 2030 |
| *ZFY* | 4740 | 0.00000199 | 0.000783 | 1930 |
| *KDM5D* | 15800 | 0.00000686 | 0.00175 | 1920 |
| *TTTY15* | 2560 | 0.0000111 | 0.00236 | 1760 |
| *UTY* | 22700 | 0.000000127 | 0.000115 | 1330 |
| *KALP* | 250 | 0.00000474 | 0.00131 | 1290 |
| *DDX3Y* | 18000 | 0.000000651 | 0.00042 | 1210 |
| *TTTY14* | 138 | 0.0000053 | 0.00141 | 543 |
| *PRKY* | 11100 | 0.00000127 | 0.000652 | 419 |
| *NLGN4Y* | 209 | 0.0000059 | 0.00153 | 309 |
| *UTS2* | 914 | 0.000667 | 0.0272 | 8.76 |
| *LOH12CR2* | 157 | 0.00000134 | 0.000657 | 6.17 |
| *AL049542.1* | 199 | 0.000427 | 0.0218 | 4.31 |
| *PDZK1* | 152 | 0.000686 | 0.0276 | 3.68 |
| *PAPPA2* | 145 | 0.00188 | 0.0456 | 3.59 |
| *CTD-2270L9.4* | 126 | 0.000211 | 0.0148 | 3.45 |
| *PPARGC1A* | 127 | 0.00136 | 0.0388 | 3.28 |
| *C9orf129* | 129 | 0.0000653 | 0.00758 | 3.07 |
| *TACSTD2* | 333 | 0.000298 | 0.0175 | 3.06 |
| *H19* | 591 | 0.0015 | 0.0404 | 3.03 |
| *C1orf127* | 136 | 0.000423 | 0.0217 | 2.86 |
| *CLEC4G* | 141 | 0.0012 | 0.0372 | 2.84 |
| *RP13-516M14.2* | 219 | 0.00101 | 0.0336 | 2.75 |
| *MDK* | 324 | 0.000561 | 0.025 | 2.66 |
| *CDC42EP1* | 954 | 0.0004 | 0.0211 | 2.63 |
| *SLC6A12* | 1400 | 0.00173 | 0.0441 | 2.61 |

| | | | | |
|---|---|---|---|---|
| *PSMC1P9* | 173 | 0.00121 | 0.0373 | 2.57 |
| *RPS2P4* | 259 | 0.00109 | 0.0349 | 2.55 |
| *RN7SL3* | 264000 | 0.0000882 | 0.00914 | 2.45 |
| *PTPRU* | 178 | 0.00156 | 0.0415 | 2.41 |
| *MBLAC1* | 428 | 0.000568 | 0.025 | 2.37 |
| *RP11-973H7.4* | 239 | 0.000563 | 0.025 | 2.33 |
| *GSTM5* | 124 | 0.000268 | 0.0165 | 2.31 |
| *ALOX15B* | 259 | 0.000834 | 0.0306 | 2.28 |
| *CEBPA* | 10500 | 0.000306 | 0.0176 | 2.20 |
| *RN7SL17P* | 343 | 0.00219 | 0.0495 | 2.19 |
| *CHST13* | 1160 | 0.000328 | 0.0185 | 2.19 |
| *RP5-1159O4.2* | 130 | 0.00146 | 0.0401 | 2.16 |
| *HBQ1* | 8400 | 0.0000937 | 0.00945 | 2.14 |
| *MMP23A* | 215 | 0.000228 | 0.0154 | 2.11 |
| *FAM27E3* | 347 | 0.00158 | 0.0415 | 2.11 |
| *RP11-783K16.14* | 139 | 0.00142 | 0.0396 | 2.10 |
| *ABHD17AP1* | 1950 | 0.00152 | 0.0408 | 2.09 |
| *AP001412.1* | 178 | 0.000808 | 0.0301 | 2.08 |
| *RP11-715J22.2* | 278 | 0.000168 | 0.013 | 2.08 |
| *RPS28* | 7560 | 0.00185 | 0.0454 | 2.08 |
| *LCNL1* | 184 | 0.000128 | 0.0115 | 2.07 |
| *HIC1* | 1430 | 0.0015 | 0.0404 | 2.05 |
| *C20orf144* | 140 | 0.00112 | 0.0353 | 2.03 |
| *ASCL2* | 3960 | 0.000022 | 0.00379 | 2.02 |
| *ARHGAP23* | 353 | 0.000857 | 0.0309 | 2.02 |
| *ECHDC3* | 1730 | 0.0000457 | 0.00647 | 2.02 |
| *LINC00266-1* | 251 | 0.000392 | 0.0208 | 2.01 |

| | | | | |
|---|---|---|---|---|
| *PTPRN2* | 4270 | 0.0000806 | 0.00866 | 2.01 |
| *CACNA1D* | 1110 | 0.000791 | 0.0298 | 2.00 |
| *ANXA2P1* | 235 | 0.000421 | 0.0217 | 2.00 |
| *GAPDHP70* | 647 | 0.00118 | 0.0367 | -2.03 |
| *TMSB4XP6* | 46500 | 0.000000000026 | 0.0000003 | -2.04 |
| *RP11-34P1.2* | 336 | 0.00139 | 0.0392 | -2.04 |
| *Y_RNA* | 83600 | 0.00102 | 0.0338 | -2.05 |
| *RNU2-6P* | 6220 | 0.00208 | 0.0482 | -2.07 |
| *RP11-549L6.2* | 429 | 0.000702 | 0.0278 | -2.07 |
| *HSPA8P5* | 1080 | 0.00000137 | 0.000657 | -2.07 |
| *SNORD13* | 52600 | 0.000581 | 0.025 | -2.10 |
| *SNORA37* | 4150 | 0.000519 | 0.0242 | -2.12 |
| *BLOC1S5-TXNDC5* | 7250 | 0.00177 | 0.0446 | -2.14 |
| *HEPH* | 538 | 0.00041 | 0.0214 | -2.15 |
| *RP5-1009N12.1* | 616 | 0.000805 | 0.0301 | -2.18 |
| *LRRN3* | 7920 | 0.000977 | 0.0328 | -2.21 |
| *RNY3* | 96600 | 0.000542 | 0.0247 | -2.24 |
| *RP11-317N8.4* | 500 | 0.000371 | 0.02 | -2.29 |
| *LRRC37A17P* | 3570 | 0.00222 | 0.0496 | -2.30 |
| *RNU2-59P* | 51200 | 0.000956 | 0.0326 | -2.34 |
| *snoU13* | 932 | 0.00000231 | 0.000888 | -2.36 |
| *RNU5E-1* | 3530 | 0.00168 | 0.0433 | -2.39 |
| *SNORA26* | 5130 | 0.000149 | 0.0124 | -2.42 |
| *RP11-333E13.2* | 1160 | 0.000000000996 | 0.000005 | -2.48 |
| *SNORA5C* | 855 | 0.0000632 | 0.00743 | -2.48 |
| *RP11-219B17.1* | 655 | 0.00000294 | 0.000972 | -2.51 |
| *CCT4P2* | 374 | 0.00139 | 0.0392 | -2.54 |

| | | | | |
|---|---|---|---|---|
| *RPL7AP31* | 1640 | 0.00000889 | 0.00203 | -2.55 |
| *RPS4XP3* | 1620 | 0.00000784 | 0.00186 | -2.58 |
| *CTA-276O3.4* | 375 | 0.00141 | 0.0394 | -2.60 |
| *RP11-552M11.4* | 1010 | 0.00127 | 0.0379 | -2.63 |
| *MTND5P5* | 680 | 0.000901 | 0.0317 | -2.66 |
| *RASSF6* | 1300 | 0.000695 | 0.0277 | -2.71 |
| *HIGD1AP14* | 366 | 0.000853 | 0.0309 | -2.72 |
| *RP5-850E9.3* | 3780 | 0.000925 | 0.032 | -2.74 |
| *SNORA45* | 1320 | 0.000261 | 0.0162 | -2.75 |
| *MAP7D2* | 430 | 0.000415 | 0.0215 | -2.89 |
| *SNORD14B* | 1380 | 0.000249 | 0.0159 | -2.90 |
| *SNORD46* | 2750 | 0.000156 | 0.0128 | -2.95 |
| *MTRNR2L9* | 36500 | 0.00000000474 | 0.0000095 | -3.01 |
| *TRBV5-5* | 414 | 0.000111 | 0.0106 | -3.03 |
| *SNORD45A* | 2210 | 0.000284 | 0.0169 | -3.12 |
| *XIST* | 723000 | 0.00000164 | 0.000699 | -3.15 |
| *RNU4ATAC* | 5250 | 0.0000227 | 0.00388 | -3.22 |
| *AC068533.7* | 623 | 0.00157 | 0.0415 | -3.23 |
| *SNORD60* | 868 | 0.00000141 | 0.000657 | -3.23 |
| *SNORD104* | 2040 | 0.000138 | 0.0119 | -3.33 |
| *GPR15* | 7120 | 0.0000015 | 0.000665 | -3.38 |
| *SNORD33* | 28200 | 0.0000000382 | 0.0000637 | -3.44 |
| *SNORD12* | 974 | 0.0000234 | 0.00391 | -3.59 |
| *AP000295.9* | 2670 | 0.00010 | 0.00994 | -3.66 |
| *SNORD118* | 15500 | 0.0000931 | 0.00945 | -3.70 |
| *SNORA3* | 1800 | 0.0000232 | 0.00391 | -3.78 |
| *AC013269.1* | 2390 | 0.000000207 | 0.000172 | -3.81 |
| *RNY5* | 1210 | 0.000132 | 0.0116 | -3.84 |

| | | | | |
|---|---|---|---|---|
| *Z82188.1* | 5550 | 0.000465 | 0.023 | -3.99 |
| *AL162497.1* | 601 | 0.000526 | 0.0243 | -4.02 |
| *SNORD35A* | 1130 | 0.00003 | 0.00458 | -4.04 |
| *SNORD100* | 245 | 0.00177 | 0.0446 | -4.22E+00 |
| *SNORD14D* | 3280 | 0.0000561 | 0.00699 | -4.27E+00 |
| *MT-TV* | 950 | 0.000161 | 0.0129 | -4.28E+00 |
| *MT-TM* | 845 | 0.000000563 | 0.000375 | -5.12E+00 |
| *SNORD71* | 286 | 0.00127 | 0.0378 | -5.17E+00 |
| *SNORD32A* | 2330 | 0.0000218 | 0.003790 | -6.47E+00 |
| *RP11-138I1.4* | 1150 | 0.000624 | 0.026 | -6.53E+00 |
| *SNORD105B* | 700 | 0.000000000000102 | 0.000000000205 | -6.73E+00 |
| *SNORD14E* | 164 | 0.0000102 | 0.00224 | -6.98E+00 |
| *SNORD56* | 408 | 0.000258 | 0.0161 | -9.28E+00 |
| *SNORD55* | 444 | 0.0000031 | 0.000984 | -1.09E+01 |
| *SNORD95* | 376 | 0.000000117 | 0.000115 | -1.11E+01 |

Genes are arranged in order of fold change. *PPARGC1A* is also known as *PGC-1α*. FDR, false discovery rate

### 3.3.4 Hierarchical Clustering

Hierarchical clustering was performed on the filtered gene list of 132 differentially expressed genes which generated a dendogram/heatmap. The standardized level of expression of each gene in each sample compared to the mean expression level of that gene across all samples was represented using the red-green colour scale shown on the key with a range from -1.63 to 5.49. Clustering did not use disease group i.e PD and controls as a variable however, cases and controls were clustered together. One case, SR033 clustered with controls in the heatmap however the expression was similar to that of the other cases.

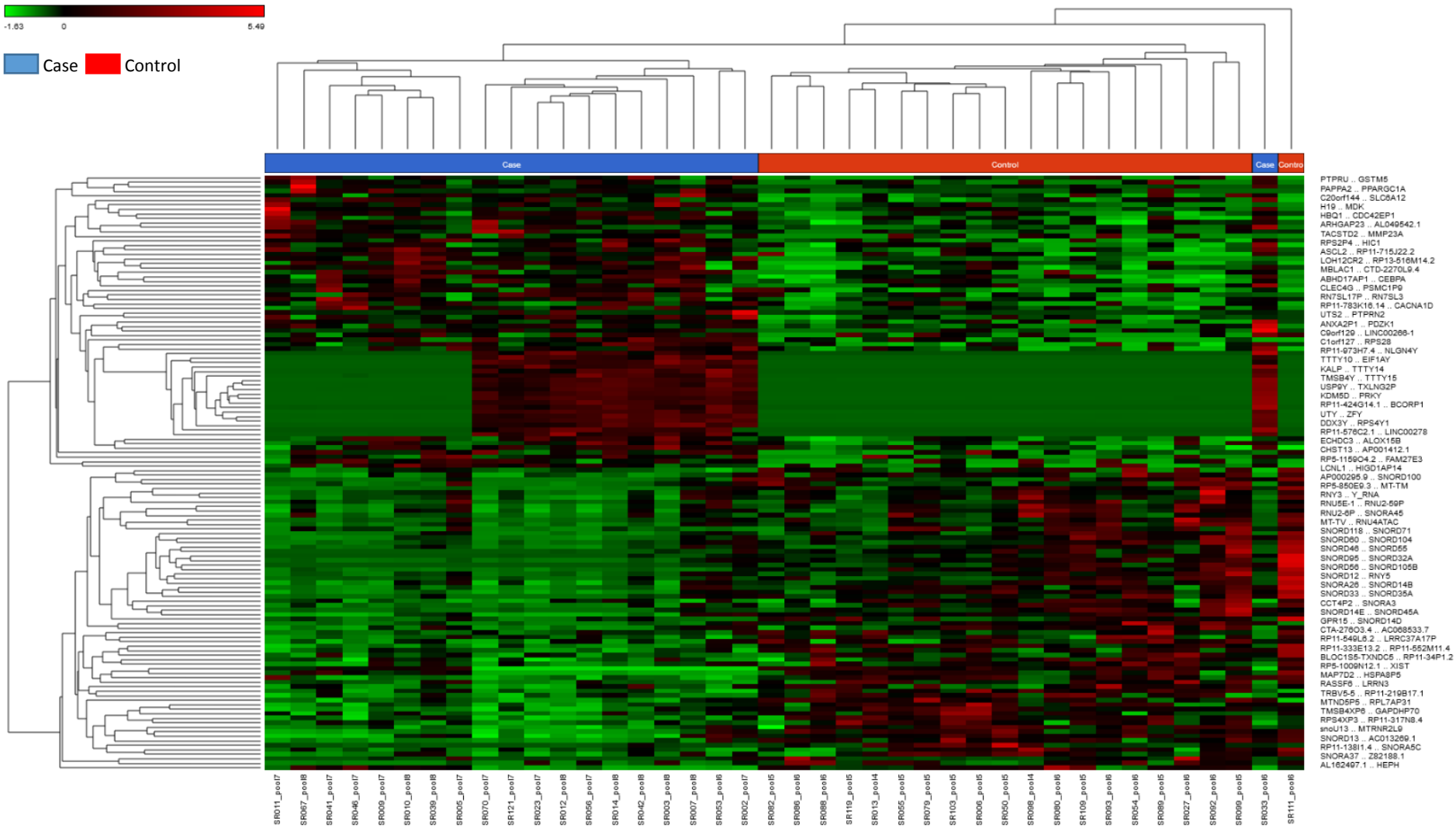



**Figure 3.7: Cluster heatmap illustrating differential gene expression between PD patient and control samples.** Green represents down-regulated genes and red represents up-regulated genes.

## 3.4 Enrichment Analysis

### 3.4.1 Gene Ontology

Genes were classified into functional groups according to Gene Ontology Annotation tools. No specific pathways were significant for the candidate gene list (N=132) for biological process, molecular function or cellular component.

### 3.4.2 Functional Annotation Clustering

Enrichment analysis was performed using DAVID Bioinformatics Database on the filtered 132 gene list for identification of common shared pathways between differentially expressed genes and this produced eight annotation clusters. Genes were grouped based on functional similarity. Two annotation clusters were of significance with an enrichment score greater than 1.3 which is equivalent to non-log scale of 0.05 (Table 3.6). These genes were grouped as they had similar functions such as iron binding and oxidoreducatse activity.

**Table 3.6: Functional annotation clustering of 132 candidate genes using DAVID Bioinformatics database**

| Annotation cluster | Keywords | Enrichment score | Count | p-value | Benjamini |
|---|---|---|---|---|---|
| 1 | Iron | 1.83 | 5 | 0.0037 | 0.37 |
| | Iron ion binding | | 5 | 0.0066 | 0.62 |
| | oxidoreductase | | 4 | 0.13 | 0.97 |
| 2 | Dioxygenase | 1.69 | 3 | 0.0096 | 0.45 |
| | Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen | | 3 | 0.011 | 0.57 |
| | Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen | | 3 | 0.012 | 0.44 |
| | oxidoreductase | | 4 | 0.13 | 0.97 |

Benjamini is a correction for multiple testing and therefore the adjusted p-value

### 3.4.3 Ingenuity Pathway Analysis (IPA)

To identify significantly dysregulated canonical pathways and generate gene networks in the sample set, the 132 genes were analysed using the IPA 17.1 software (IPA, Ingenuity Systems, www.ingenuity.com) and of these 106 were mapped to the IPA knowledge database. Canonical pathways are constructed using IPA pathway diagrams for well established signaling and metabolic pathways, information from Kyoto encyclopedia of genes and genomes (KEGG) and LIGAND are also included for canonical pathway identification. These pathways show key biological and molecular roles of proteins at the cellular level. Molecule shapes indicate what protein family this molecule belongs to such as an enzyme, cytokine etc and the relationship between molecules are indicated by the different arrows (Appendix XVI).

Five significantly deregulated pathways were identified based on p-values (Table 3.7) which included: regulation of eukaryotic initiation factor 4 (eIF4) and p70S6K signaling, eukaryotic initiation factor 2 (EIF2) signaling, lipopolysaccharide/ interleukin 1 (LPS/IL-1) mediated inhibition of retinoid X receptors (RXR) function, xenobiotic metabolism signaling, and maturity onset diabetes of young (MODY) signaling. In Figure 3.8, significant canonical pathways are plotted according to p-value indicated by the bar and the number of differentially expressed genes divided by the total number of genes for that pathway depicted by the yellow square.

**Table 3.7: Five significant canonical pathways found in whole blood from PD patients using Ingenuity pathway analysis.**

| Canonical pathway | p-value |
|---|---|
| Regulation of eIF4 and p70S6K signaling | 0.00519 |
| EIF2 signaling | 0.00979 |
| LPS/IL-1 mediated inhibition of RXR function | 0.0160 |
| Xenobiotic metabolism signaling | 0.0277 |
| Metabolism signaling, maturity onset diabetes of young (MODY) signaling | 0.0492 |

EIF2, Eukaryotic Initiation Factor 2; eIF4, Eukaryotic Initiation Factor-4; IL-1, interleukin-1; LPS, lipopolysaccharide; RXR, Retinoid X receptor. IPA, Ingenuity Systems, www.ingenuity.com

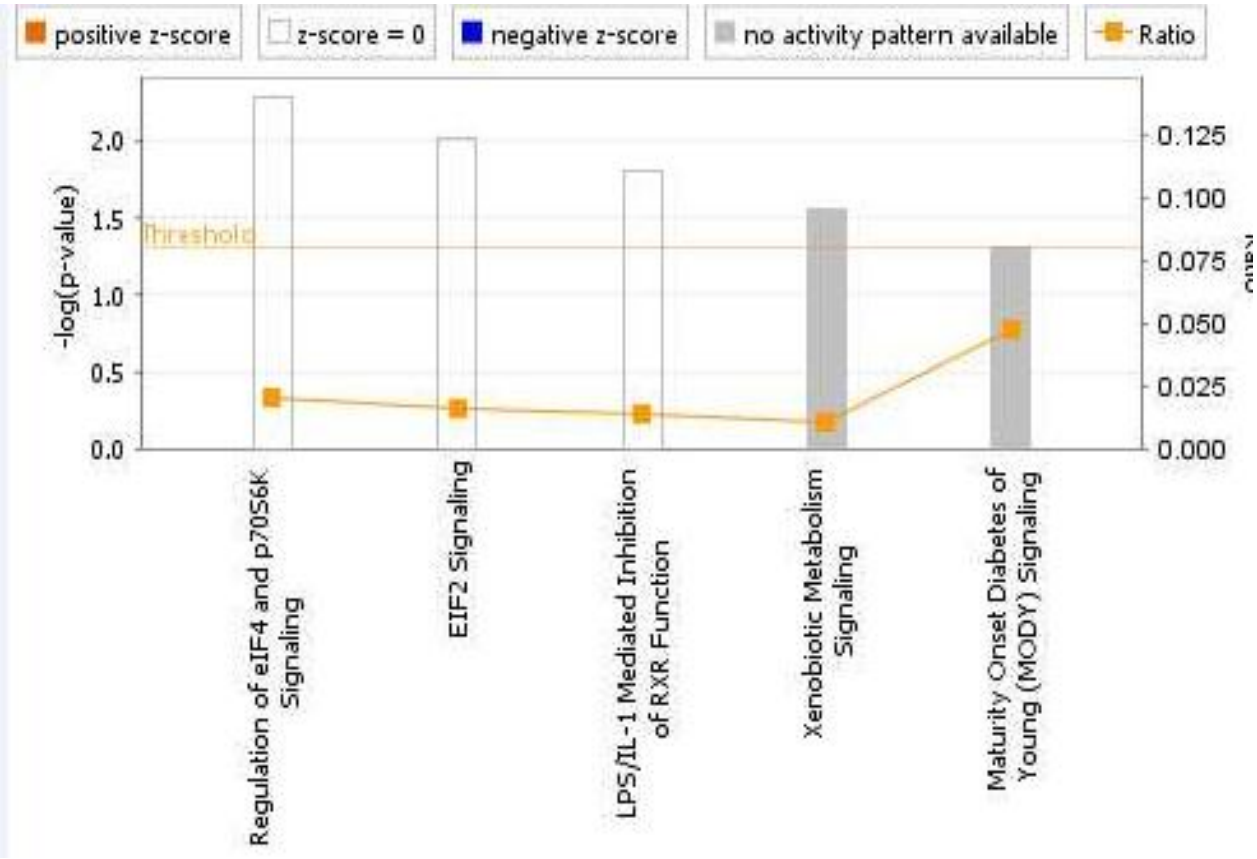



**Figure 3.8: The five canonical pathways found to be dysregulated between Parkinson's disease patients and controls using Ingenuity pathway analysis.** For each canonical pathway, the bar represents the p-value and the square represents the number of differentially expressed genes divided by the total number of genes for that pathway.

IPA also generated gene networks from the list of 132 genes, and the networks containing the highest number of genes included those involved in cellular movement, cell to cell signaling and interaction, drug metabolism (Figure 3.9), endocrine system development and function, molecular transport and small molecule biochemistry (Figure 3.10) as well as cellular development, cellular growth and proliferation and hematological system development and function (Figure 3.11). These molecular networks are generated using algorithms to describe potential molecular interactions in this experimental system between the 132 dysregulated genes. Five toxicology (tox) lists of significance were also highlighted (Table 3.8). The toxicology lists reveal biological mechanisms related to toxicity which are either molecular, cellular or biochemical for identification of toxic compounds related to the disease.Therefore pathways or biological mechanisms that are postulated to be adversely affected by the 132 candidate genes.

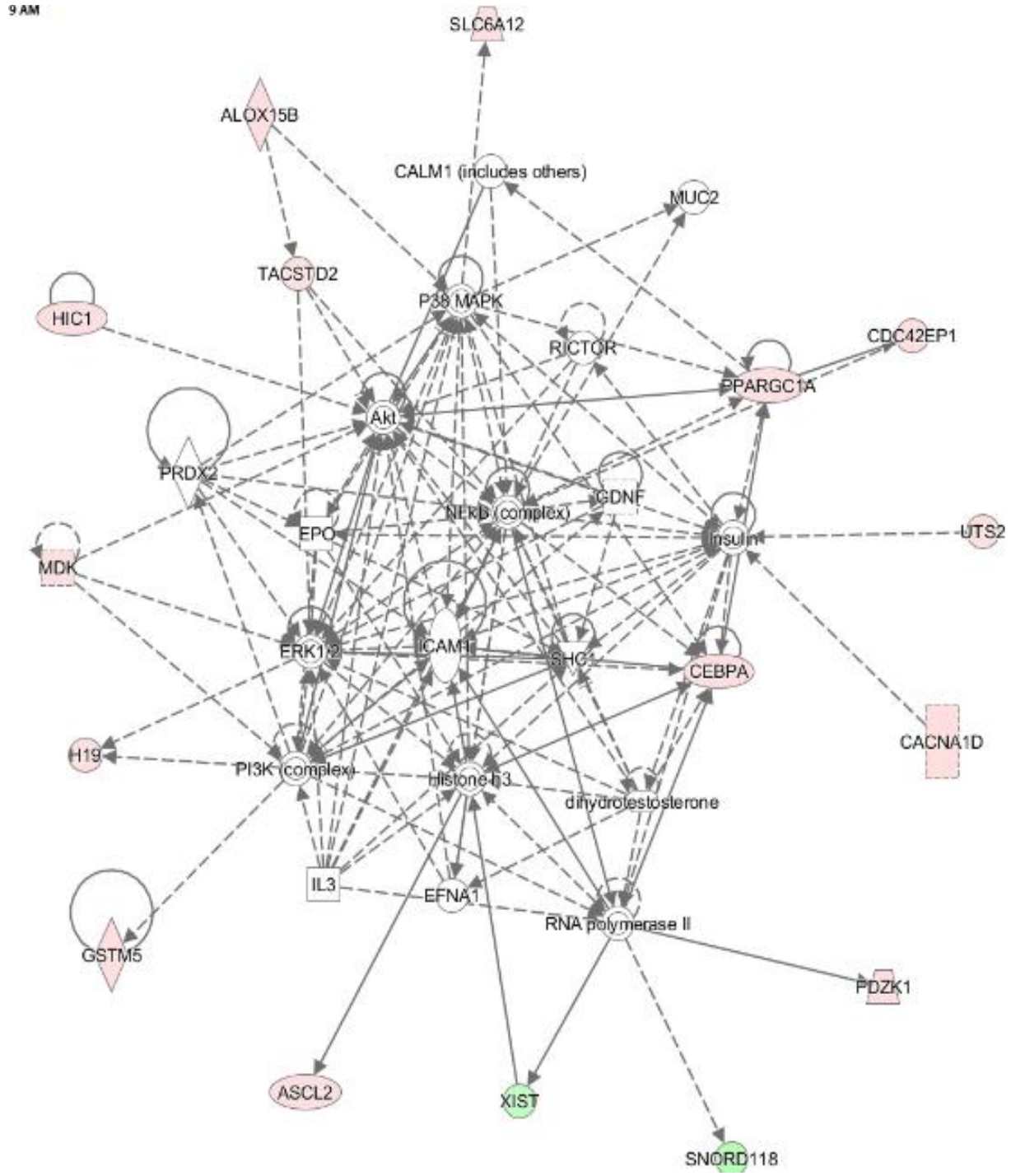



**Figure 3.9: Gene expression network identified from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using ingenuity pathway analysis (IPA).** Genes present in the representative network were involved in cellular movement, cell to cell signaling and interaction and drug metabolism

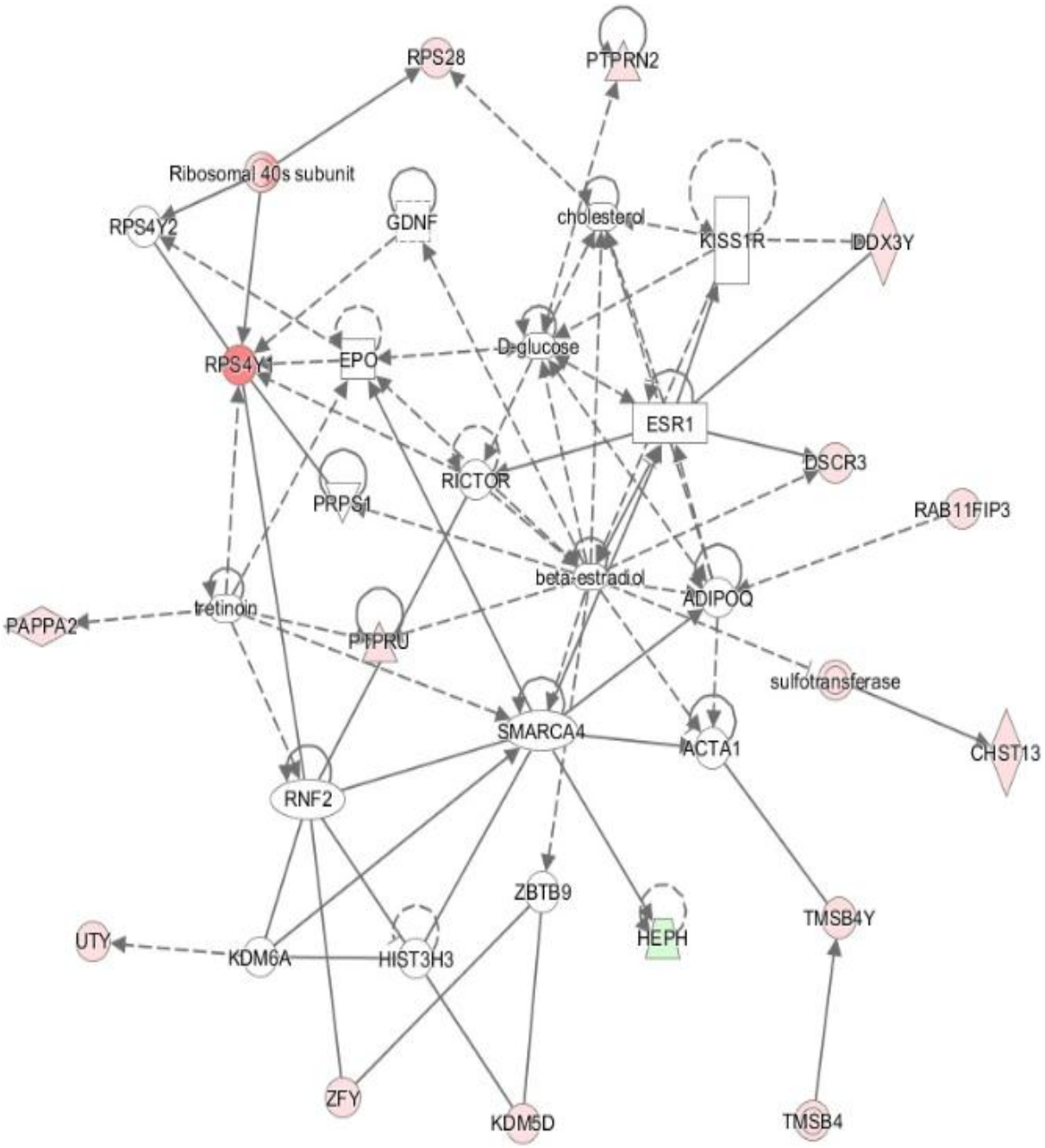



**Figure 3.10: Gene expression network from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using ingenuity pathway analysis (IPA).** Genes present in the representative network were involved in endocrine system development and function, molecular transport and small molecule biochemistry.

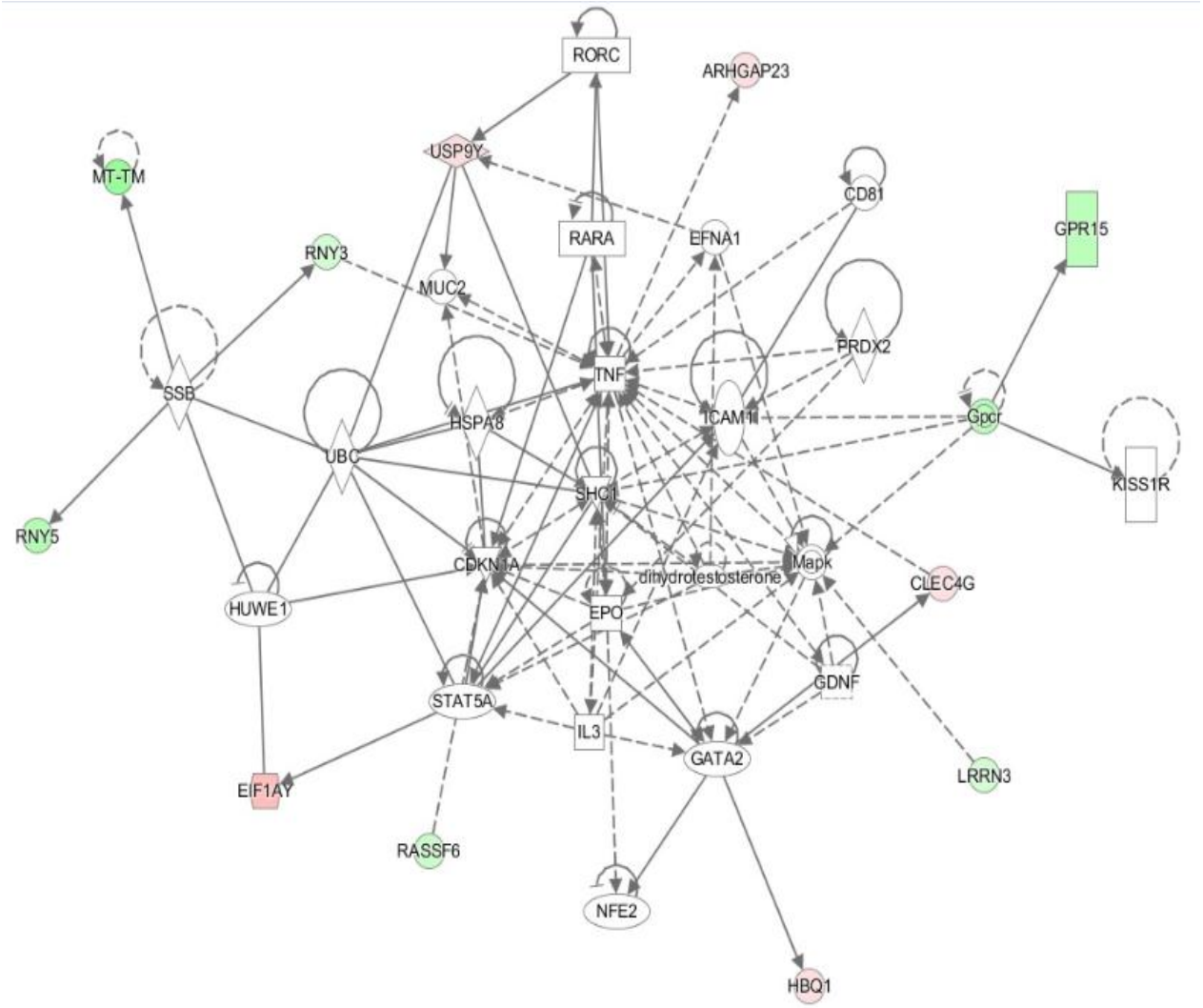



**Figure 3.11: Gene expression network from 132 genes found to be significantly dysregulated between Parkinson's diease patients and controls using Ingenuity Pathway Analysis (IPA).** Genes present in the representative network were involved in cellular development, cellular growth and proliferation and hematological system development and function.

**Table 3.8 Five significant tox lists found in whole blood from PD patients using Ingenuity pathway analysis.**

| Tox list name | p-value |
|---|---|
| Biogenesis of mitochondria | 0.00104 |
| Renal inorganic phosphate homeostasis (mouse) | 0.0143 |
| Increases liver steatosis | 0.0202 |
| LPS/IL-1 mediated inhibition of RXR function | 0.0226 |
| Increases bradycardia | 0.0354 |

LPS, lipopolysaccharide; IL-1, interleukin-1

## 3.5 Gene Prioritization

For further downstream analysis of potential genes involved in PD pathology, a comprehensive search of genes identified in previous gene expression studies on PD was performed. The gene lists for each published study was compiled and this resulted in 4,199 candidate genes from studies on blood and 12,117 genes from studies on brain (Appendix I and II, Tables 1 and 2). The 132 candidate genes from the present study were then compared to these tables. Genes were prioritized if differentially expressed in the same direction (i.e up or down-regulated compared to controls) in at least one study using blood and in at least one study using brain tissue as well as in the present study.

A total of four genes met these criteria (Table 3.9) and these included: Taxilin gamma pseudogene, Y-linked (*TXLNG2P*), Ubiquitin specific peptidase 9, Y-linked (*USP9Y*), Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked (*UTY*) and CCAAT/enhancer binding protein alpha (*CEBPA*).

**Table 3.9: Genes prioritized from candidate gene list based on a literature search identifying genes in the same direction (up/ down) of differential expression.**

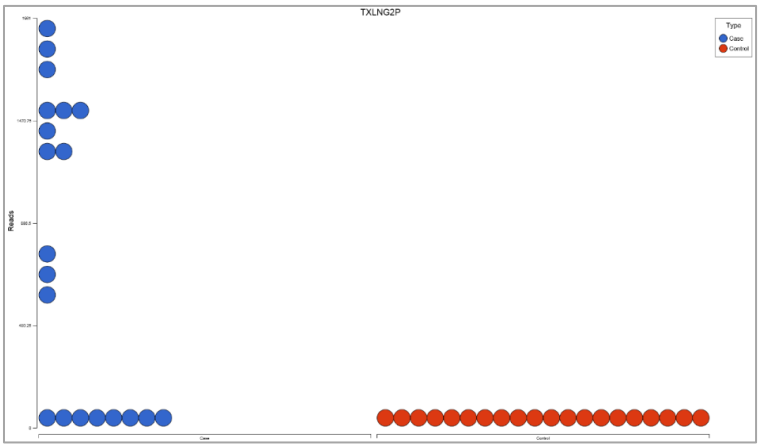
| Gene Symbol | The present study | Published studies using blood | | Published studies using brain tissue | | |
|---|---|---|---|---|---|---|
| | Up/ Down regulation | Reference | Up/ Down regulation | Reference | Brain region | Up/ Down regulation |
| ***TXLNG2P*** | Up | Santiago et al, 2015 | Up | Riley et al, 2014 | CTX<br>SN | Up<br>Up |
| ***USP9Y*** | Up | Soreq et al, 2008 | Up<br>Up | Riley et al, 2014 | CTX<br>SN | Up<br>Up |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Str | Up |
| ***UTY*** | Up | Santiago et al, 2015 | Up | Riley et al, 2014 | CTX<br>SN | Up<br>Up |
| ***CEBPA*** | Up | Santiago et al, 2015 | Up | Riley et al, 2014 | SN | Up |

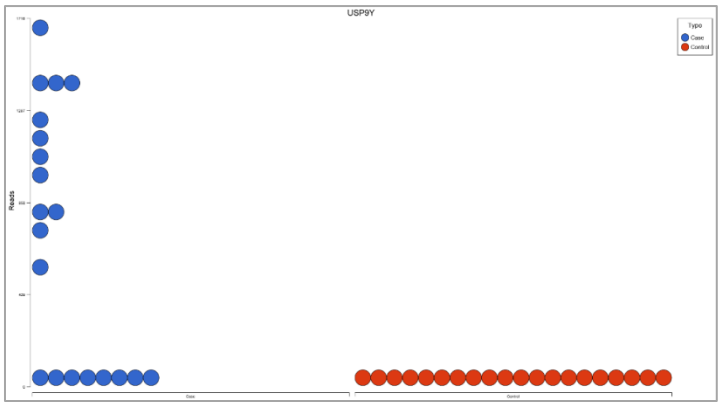
*CEBPA*, CCAAT/enhancer binding protein alpha; CTX, cortex; SN, substantia nigra; Str, striatum; *TXLNG2P*, Taxilin gamma pseudogene, Y-linked; *USP9Y*, Ubiquitin specific peptidase 9, Y-linked; *UTY*, Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked.

Cases and controls were also grouped according to the number of reads. The generated reports for each of the candidate genes are shown in Figure 3.12. Each circle represents a different sample. Cases are displayed as blue whereas controls are displayed as red. According to the number of reads cases and controls were distinctly clustered showing no overlap for each of the candidate genes. Some samples had a much higher read number indicated by the position on the y-axis.

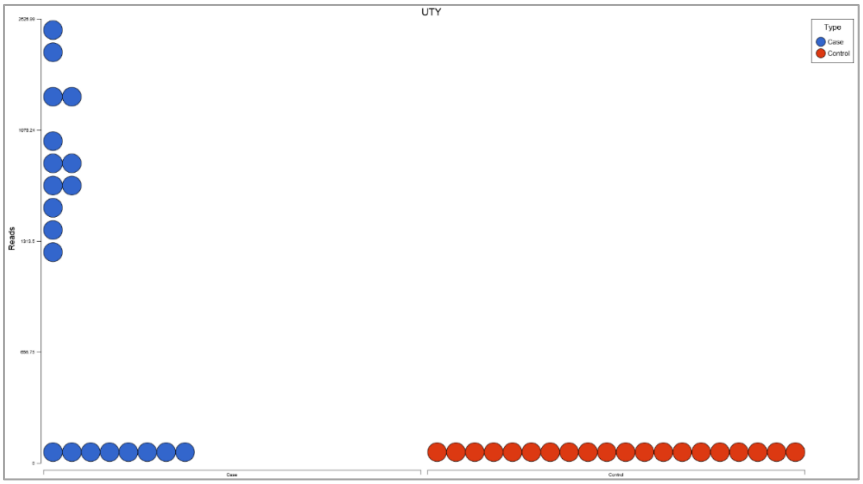
**A**




**B**



**C**




**D**

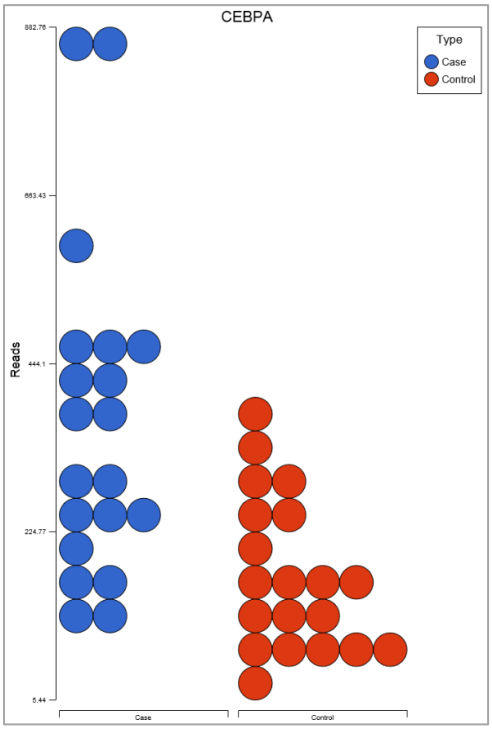



**Figure 3.12: Clustering of cases and controls according to reads using Partek Flow for each prioritized candidate gene.** A) *TXLN2GP*, B) *USP9Y*, C) *UTY* and D) *CEBPA*

A literature search was performed using the National Center for Biotechnology Information (NCBI) (http://www ncbi.nlm nih.gov/RefSeq/), Online Mendelian Inheritamce in Man (OMIM) (http://www.omim.org/) and GeneCards® (http://www.genecards.org/) databases to identify the function of the four prioritized candidate genes to determine which of these genes were plausible candidate genes for further study (Table 3.10).

**Table 3.10: Function of four prioritized candidate genes**

| Gene symbol | Gene name | Entrez ID | OMIM | Function |
|---|---|---|---|---|
| ***TXLNG2P*** | Taxilin gamma pseudogene, Y-linked | 246126 | Not assigned | • Pseudo-gene and is Y-linked. |
| ***USP9Y*** | Ubiquitin specific peptidase 9, Y-linked | 8287 | 400005 | • Member of the peptidase C19 family<br>• Encodes a protein that is similar to ubiquitin-specific proteases<br>• Expression in the testes and association with Sertoli cell-only syndrome (Sargent et al., 1999) |
| ***UTY*** | Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked | 7404 | 400009 | • Encodes a protein containing tetratricopeptide repeats which are thought to be involved in protein-protein interactions<br>• Processing of ubiquitin precursors and of ubiquitinated proteins<br>• Component of the *TGF-β/*BMP signaling cascade<br>• Regulation of prostrate differentiation and association with prostrate cancer (Dutta et al., 2016) |
| ***CEBPA*** | CCAAT/enhancer binding protein alpha | 1050 | 116897 | • Intronless gene which encodes a basic leucine zipper (bZIP) domain and recognizes the CCAAT motif in the promoters of target genes<br>• Modulates leptin expression and treatment for human obesity (Miller et al., 1996)<br>• Regulates gluconeogenesis and lipogenesis in the liver<br>• Inhibition of cyclin dependant kinases (Wang et al., 2001)<br>• Down regulates cancer cell growth (Reddy et al., 2009)<br>• Mutations cause acute myeloid leukemia (Smith et al., 2004) |

BMP, bone morphogenetic protein; *CEBPA*, CCAAT/enhancer binding protein alpha; NCBI, National Center for Biotechnology Information; *TXLNG2P*, Taxilin gamma pseudogene, Y-linked; *TGF-beta*, Transforming growth factor beta 1; *USP9Y*, Ubiquitin specific peptidase 9, Y-linked; *UTY*, Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked.

However, given the gender discrepancy between the cases and controls (all the controls were female), and the fact that *TXLNG2P*, *USP9Y* and *UTY* are all Y-linked genes, it was decided to not perform further validation experiments on these three genes.

## 3.6 Verification of *CEPBA* gene

The *CEBPA* gene (OMIM 116897) was selected for verification using qPCR. Cq values for both patient and control samples were generated using the Lightcycler® 96 software (Roche Diagnostics, Indianapolis, USA). All samples were run in triplicate and all Cq values are listed in Appendix XVII, Table 9. One patient sample, SR039 showed large technical variation in triplicate Cq values for both the target gene and three reference genes and was therefore removed from further analysis. Hence, qPCR data analysis of *CEBPA* gene expression was performed on 19 patient and 20 control samples.

Raw Cq values were analysed using REST-2009 © software (Qiagen, Hilden, Germany) (Pfaffl et al., 2002). *CEBPA* was calculated to be down-regulated in cases compared to controls by a fold change of 0.579 with a p-value of 0.010, when *CEBPA* expression was normalized to all three reference genes (Table 3.11)

**Table 3.11: Comparison of *CEBPA* expression to three reference genes *HBB*, *GAPDH* and *ACTB* using REST-2009© software**

| Gene | Type | Expression | Std. Error | 95% C.I. | p-value | Up/down regulation |
|---|---|---|---|---|---|---|
| *HBB* | Reference | 1.151 | | | | |
| *GAPDH* | Reference | 0.726 | | | | |
| *ACTB* | Reference | 1.196 | | | | |
| *CEBPA* | Target | 0.579 | 0.124 - 2.605 | 0.030 - 11.237 | 0.010 | Down |

*ACTB*, β-actin; CI, Confidence interval; *CEBPA*, CCAAT/enhancer binding protein alpha; *GAPDH*, Glyceraldehyde 3-phosphate dehydrogenase; *HBB*, hemoglobin subunit beta; Std, Standard

Non-normalized results for all three reference genes and the target gene were also generated by the REST-2009 © software. Notably, in this analysis, *GAPDH* was shown to be significantly down-regulated in patients compared to controls with a fold change of 0.591 and a p-value of 0.038 (Table 3.12).

**Table 3.12: Non-normalized results of *CEBPA*, *HBB*, *GAPDH* and *ACTB* expression using REST-2009 © software**

| Gene | Type | Expression | Std. Error | 95% C.I. | p-value) | Up/down regulation |
|---|---|---|---|---|---|---|
| *HBB* | Reference | 0.937 | 0.165 - 5.247 | 0.047 - 37.033 | 0.766 | |
| *GAPDH* | Reference | 0.591 | 0.084 - 4.004 | 0.015 - 11.634 | 0.038 | Down |
| *ACTB* | Reference | 0.973 | 0.227 - 5.464 | 0.051 - 25.817 | 0.900 | |
| *CEBPA* | Target | 0.471 | 0.032 - 4.000 | 0.006 - 22.951 | 0.014 | Down |

*ACTB*, β-actin; CI, Confidence interval; *CEBPA*, CCAAT/enhancer binding protein alpha; *GAPDH*, Glyceraldehyde 3-phosphate dehydrogenase; *HBB*, hemoglobin subunit beta; Std, Standard

Therefore, it was decided that *GAPDH* was not a suitable reference gene for this dataset, and it was removed and the data re-analysed using only *HBB* and *ACTB* as reference genes. Upon re-analysis, *CEBPA* was again found to be significantly down-regulated in cases compared to controls by a factor of 0.493 with a p-value of 0.005 (Table 3.13).

**Table 3.13: Comparison of *CEBPA* expression to two reference genes *HBB*, and *ACTB* using REST-2009 © software**

| Gene | Type | Expression | Std. Error | 95% C.I. | p-value) | Up/down regulation |
|---|---|---|---|---|---|---|
| *HBB* | Reference | 0.981 | | | | |
| *ACTB* | Reference | 1.019 | | | | |
| *CEBPA* | Target | 0.493 | 0.084 - 3.164 | 0.016 - 13.790 | 0.005 | Down |

*ACTB*, β-actin; CI, Confidence interval; *CEBPA*, CCAAT/enhancer binding protein alpha; *HBB*, hemoglobin subunit beta; Std, Standard

And finally, for comparative purposes, the qPCR data was also analysed using the $2^{-\Delta\Delta Cq}$ method and a Microsoft Excel Spreadsheet (Vandesompele et al., 2002). Each sample was normalized to *HBB* and *ACTB*, and the normalized fold changes in *CEBPA* expression for each sample individually, is shown in Figure 3.13. This graph demonstrates visually the overall trend of lower *CEBPA* expression in patient samples compared to control samples. Cases and controls were also grouped in order to calculate the mean fold change in *CEBPA* expression between the two groups, whereas a two-tailed unpaired

Student's T-test was used to investigate statistical significance. Down-regulation of *CEBPA* was observed in cases compared to controls with a mean factor of 0.480, with a p-value of 0.003 (Figure 3.14). Hence, analysis of the qPCR data by means of the $2^{-\Delta\Delta Cq}$ method concurred with the analysis performed on REST-2009 © software. However, these findings are the exact opposite of the RNA-seq data analysis, which found up-regulation of *CEBPA* expression in PD patients compared to controls.

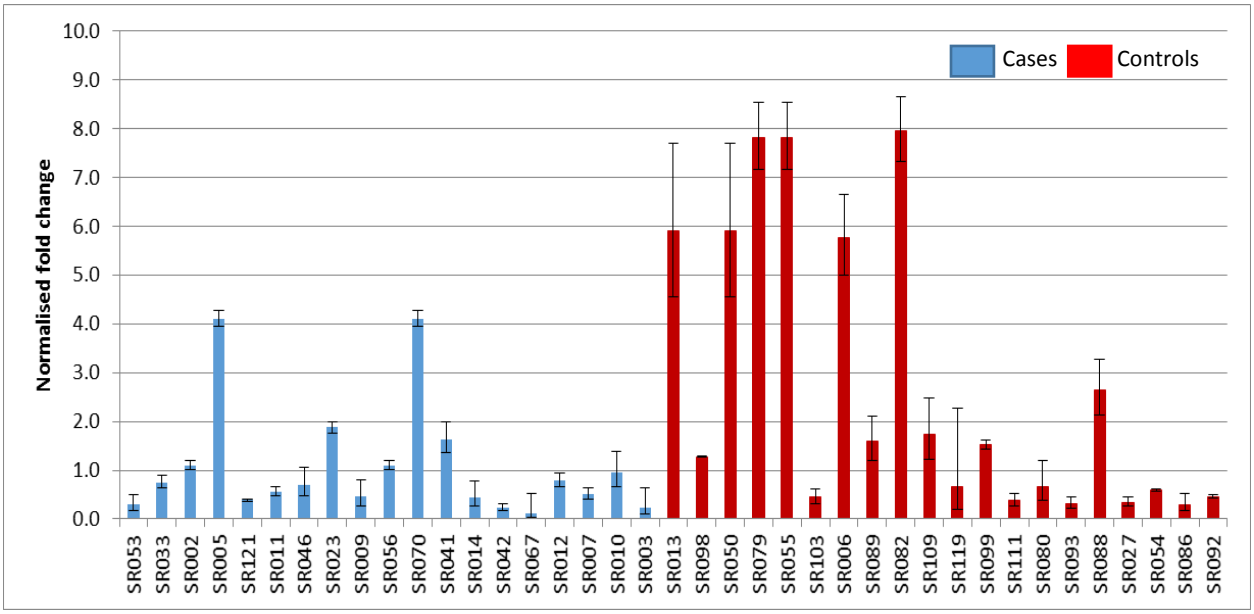



**Figure 3.13: Verification of differential gene expression of *CEBPA* by qPCR using the Lightcycler® 96**. Data were analysed using the $2^{-\Delta\Delta Cq}$ method (Vandesompele et al., 2002) and plotted as normalized fold change of the target gene *CEBPA* for each PD patient and control sample when normalized to two reference genes, *HBB* and *ACTB*. Bars depict normalized fold changes in *CEBPA* expression ± SE. PD patient samples are depicted as blue and controls depicted as red. SE, standard error

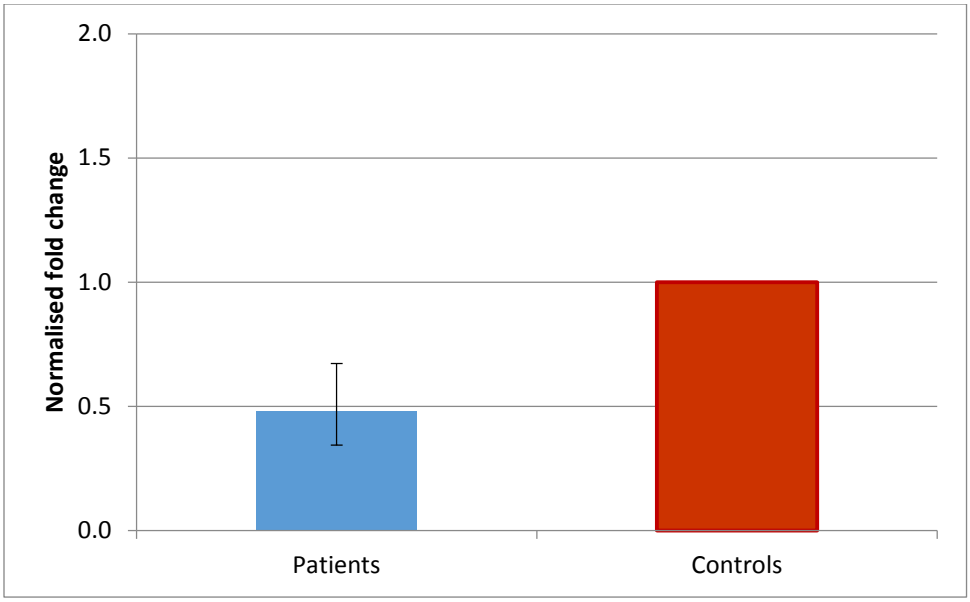



**Figure 3.14: Verification of differential gene expression of *CEBPA* by qPCR.** Data were analysed using the $2^{-\Delta\Delta Ct}$ method (Vandesompele et al., 2002) and plotted as relative expression of the target gene *CEBPA* for PD patient and control samples when normalized to two reference genes, *HBB* and *ACTB*. Bars depict normalized fold changes in *CEBPA* expression ± SE. Grouped PD patient samples are depicted as blue and grouped controls depicted as red.

# Chapter 4: Discussion

4.1 Detection of a heterozygous *PARK2* exon 2 deletion ..... 107
4.2 Quality Control of RNA and cDNA ..... 109
4.3 RNA-Seq Pipeline ..... 110
4.4 Biological pathways and enrichment analysis ..... 112
4.5 Link between *CEBPA*, *PGC1α* and PD ..... 113
4.6 Verification using qPCR ..... 118
4.7 Study limitations ..... 119
4.8 Future Work ..... 120
4.9 Conclusions ..... 122

Whole genome transcriptomic profiling offers a snapshot of potentially all of the genes that are being expressed at a specific time point and allows the relative quantification of all RNA isoforms. To date, research efforts have mainly focused on utilizing microarray for transcriptomic profiling and very few studies have used RNA-Seq specifically in the field of PD. Therefore, the present study focused on transcriptomic profiling using RNA-Seq in a subset of 40 study participants selected from the Shared Roots parent study to investigate whether significant differences in gene expression and pathways between PD patients and controls could be detected. Initially, all PD patients were screened using MLPA to exclude patients with CNV mutations in the known PD-causing genes. Following from this, RNA-Seq was performed and a total of 132 genes were found to be differentially expressed between the PD patients and controls. Upon filtering, four genes were highlighted as potential candidates and *CEBPA* was finally selected for further verification using qPCR in all of the patient and control samples. Notably, this gene was found to be up-regulated in patients using RNA-Seq but down-regulated using qPCR.

## 4.1 Detection of a heterozygous *PARK2* exon 2 deletion

CNV mutations are known to play a significant role in the development of Mendelian forms of PD (Pankratz et al., 2011). All of the PD patients were screened for CNV mutations in the known PD-genes as well as two point mutations (*SNCA A30P* and *LRRK2 G2019S* ) using MLPA. An exonic rearrangement in the form of a heterozygous *PARK2* exon 2 deletion was found in one male patient, SR023 (Figure 3.1). Verification using qPCR confirmed the MLPA results that *PARK2* exon 2 in patient SR023 was only present on one allele compared to healthy controls (WT 1 and 2) in whom both *PARK2* exon 2 alleles were present (Figure 3.2).

Mutations in *PARK2* were first identified in Japanese patients with autosomal recessive juvenile parkinsonism (Kitada et al., 1998). Since then exonic rearrangements in *PARK2* have been studied by various researchers in numerous populations worldwide including the Polish (Koziorowski et al., 2010), Russian, German (Periquet et al., 2001) and North American (Pankratz et al., 2009) populations. Kilarski and colleagues performed a systematic review which included 43 articles in which *PARK2* mutations in five patients were identified (Kilarski et al., 2012). Exonic rearrangements included deletions, multiplications, single base pair substitutions and small deletions or insertions of one or several base pairs. This review included white, Asian, Latin American and other (Arabic, black African or Caribbean) ethnic populations.

Exonic rearrangements in African PD patients are understudied, and to date only a few *PARK2* CNV mutations in the South African population have been identified (Bardien et al., 2009; Haylett et al., 2012; Keyser et al., 2010; van der Merwe et al., 2016). Keyser and colleagues identified a homozygous deletion of *PARK2* exon 3 and 4 in one patient and a heterozygous duplication of *PARK2* exon 2 and exon 3 in another (Keyser et al., 2010). A second study which expanded on previous work of Keyser *et al* identified two PD mixed ancestry patients with *PARK2* mutations which included, a heterozygous deletion of *PARK2* exon 4 and a heterozygous deletion of *PARK2* exon 3 and 4 (Haylett et al., 2012). More recently Van der Merwe and colleagues screened a total of 210 PD patients and a heterozygous *PARK2* exon 4 deletion was identified in one male patient (van der Merwe et al., 2016). However, a second mutation in parkin was not identified in this patient through Sanger sequencing. Therefore, in total from previous studies on South African PD patients, out of 439 patients screened, only 7 *PARK2* CNV mutations were found in South African PD patients.

PD-causing mutations in *PARK2* are either homozygous or compound heterozygous. Therefore, in the present study Sanger sequencing was performed to identify whether another mutation was present in any other of the 12 exons of *PARK2* in patient SR023. However, no other pathogenic mutations were found (Figure 3.3.).

Whether a single heterozygous mutation in *PARK2* can result in PD onset or cause increased risk is unknown (Kay et al., 2010). However, several heterozygous alterations of *PARK2* have been detected in healthy individuals which included missense mutations, frameshift and gene dosage alterations (Brüggemann et al., 2009). This indicated that single heterozygous *PARK2* mutations are present in healthy individuals but whether they are at increased risk of PD development is questionable.

It is possible that this patient may harbour a second mutation in a regulatory region of *PARK2* (Morrison, 2003) that was not screened in the present study. However, the likelihood of this is low and also the phenotype in this patient is not consistent with *PARK2*-associated PD. These patients usually present with early onset (< 45 years) autosomal recessive parkinsonism, dystonia and slow progression of the disease (Lücking et al., 2000). This was not seen in individual SR023 as his age of onset was 70 years. Since no pathogenic mutations in known genes were detected, these PD patients potentially have idiopathic PD and could therefore be analysed as one group for the RNA-Seq analysis.

## 4.2 Quality Control of RNA and cDNA

Quality controls problems have significant implications for further downstream analysis and experiments and it was therefore of the utmost importance that quality control at various stages of RNA-Seq analysis was performed. Following RNA isolation, RNA concentration and purity was determined using the Agilent 2100 Bioanalyser. RIN values greater than 7 are necessary for RNA-Seq as low RIN values could indicate the presence of inhibitors or contaminants. RIN values in this study ranged from a minimum value of 7.3 to 10 (Table 3.1) and this was indicative of good quality, intact RNA with minimal degradation. A virtual gel and electropherogram were generated using the Agilent Bioanalyser software. The 18S and 28S peaks at the expected fragment sizes were visible on the virtual gel (Figure 3.4) RNA quality and purity was also checked by NXT-Dx (Gent, Belgium) on arrival of the RNA in Belgium and before cDNA library preparation was performed. A comparison of the RIN values is presented in Table 3.2. Overall, the RNA was of good quality and passed various quality control steps to be acceptable for downstream experimental purposes.

Genomic DNA contamination could have occurred during RNA isolations and therefore an in house quality control test was performed on cDNA converted from RNA using primers for *HBB*, a reference gene. Minimal gDNA contamination was found and successful cDNA conversion was achieved. If gDNA contamination was present in the RNA samples then the RNA-Seq results would prove questionable as the sequencing reads would be of cDNA as well as gDNA.

Probably the most important step in RNA-Seq was the library preparation of cDNA. It should be noted that it was the cDNA that was converted from the RNA that was actually sequenced, and therefore quality control at this stage of the experiment was of the utmost importance. Quality control of the cDNA library was performed using the Agilent 2100 HS DNA Chip, and a peak at 300 bp with a narrow distribution was observed, as was expected. All peaks should be detectable with a clear visible baseline which was observed for all 40 samples (Table 3.3). Therefore, the samples were of good quality for sequencing on the Illumina HiSeq 4000 (Illumina Inc, San Diego, USA). This sequencing system is based on sequencing by synthesis (SBS) chemistry which uses a reversible terminator based method to detect single nucleotide bases as they are incorporated into DNA template strands. This sequencing system is thought to generate highly accurate base by base sequencing and raw reads which contain few sequence-context specific errors. Therefore, this sequencing system generates reads with very few errors and high percentage base calls.

## 4.3 RNA-Seq Pipeline

RNA-Seq data analysis can be a daunting task for a biologist inexperienced with using command line, Unix and R packages. A simpler and more user friendly approach to RNA-Seq data analysis is to use a graphical user interface, some of these include BaseSpace® (https://basespace.illumina.com/home/index), Chipster (http://chipster.csc.fi/), Galaxy (galaxyproject.org/) and Partek® Flow® (http://www.partek.com/partekflow). Partek® Flow® was utilized in this study as it was readily available through the core facility at Stellenbosch University (CAF) Also, this software package hosts a variety of tools allowing processing of raw RNA-Seq reads to the identification of differentially expressed genes and offers the option of building or downloading additional bioinformatic tools to produce a customized data analysis pipeline. The RNA-Seq pipeline used for investigation of differential gene expression in this study is represented in Figure 3.6. Bioinformatic tools recommended by CAF were utilized.

New bioinformatic tools for RNA-Seq data analysis are constantly being developed and improved upon (Soneson and Delorenzi, 2013). Constant development means more precise outcomes and it should be noted that the raw data can always be re-analyzed using different pipelines, as new software becomes available. It is important to note however that the gigo (garbage in, garbage out) rule applies i.e. data analysis is highly dependent on the quality of the sequencing reads.

Pre-alignment quality control performed using FastQC revealed that the reads were of high quality, as the average quality of all reads was above 30 (Appendix XV, Table 8).The average read length across all samples was 50 bp and minimal trimming was required. Quality trimming of reads removed around 8 bp which resulted in reads of around 42 bp in length.

Accurate alignment of reads is the ultimate purpose of alignment tools however this proves challenging as reads could contain mismatches, insertions and deletions introduced through sequencing errors and genomic variations. To overcome the potential challenge of inaccurate alignment, reads were aligned using STAR RNA-Seq aligner. STAR has been shown to outperform other sequencing aligners with a higher mapping speed and alignment sensitivity (Dobin et al., 2013). Also, higher mapping accuracy, fewer spliced mappings, fewer mismatches of primary alignments and high base-wise accuracy are some advantages of the STAR aligner as this would reduce incorrect placement of reads and false exon junctions reported (Engström et al., 2013). Other aligners have been shown to have high mapping error rates and low mapping speed.

Reads were mapped to the reference genome (hg19).The read length is the number of bases sequenced and the total read length for all samples ranged from 29 M to 55 M. The mapped reads were quantified using the Partek E/M algorithm. This algorithm was designed for reconstruction of full-length transcript isoforms with the ability to reduce fragmentation (Xing et al., 2006). However, Partek E/M is not widely used and other bioinformatic tools are generally preferred in RNA-Seq studies for quantification including HTSeq and Cufflinks.

Differential gene expression was performed using the GSA algorithm which forms part of the Partek® Flow® software. The output of this algorithm was a gene list table which contained a total of 19,983 unfiltered differentially expressed genes that were present in the dataset. This is a large number of differentially expressed genes considering that there is an estimated 20,500 human protein coding genes in the human genome (Clamp et al., 2007). Some of the genes detected could be alternative splicing isoforms, pseudogenes or sequencing artifacts and therefore to investigate differentially expressed genes of relevance, the dataset was filtered further.

A total of 132 genes showed differential expression between PD patients and controls (Fold change of ≤-2 and ≥ +2, p-value of ≤ 0.01 and FDR of ≤ 0.05) (Table 3.5) Genes were clustered which generated a dendogram (Figure 3.7). Overall cases and controls were clustered separately and distinct patterns of up and down regulation of genes were observed. This was used to give an overall indication of the standardized level of expression for each gene in each sample. However, one case (SR033) clustered with the control samples. This was not completely unexpected as clustering is based on similarity of samples and due to the small sample size used this could be a false positive result which was undetected by the algorithm.

Researchers have however highlighted the limitations associated with arbitrarily setting parameter cut-offs such as selecting a specific fold change cut-off which could include false positives in the dataset (Simunovic et al., 2010; Stamper et al., 2008). It is also necessary to bear in mind that significant parameter cut-offs do not necessarily correlate with biological relevance. In other words, prioritizing genes based on the lowest p-values may not be the best way to enrich for plausible candidates.

Therefore, to prioritize candidate genes from the filtered list, a comprehensive search was performed on all literature on genome wide studies performed in blood and brain from PD patients. A total of 33 articles were included which covered 17 genome wide studies which used blood as a tissue source and 16 studies which used brain tissue. From these articles two tables were generated for identification of overlapping genes, refer to Appendix I and II, Tables 1 and 2 for the complete set of differentially expressed genes identified in blood and brain. The candidate gene list (N=132) was compared to these

tables and filtered to a candidate list based on the reasoning that the gene had to be present in both blood and brain tables and had to be expressed in the same direction. For example, if a gene was upregulated in the 132 differentially expressed genes list, the gene had to be present in a minimum of one blood and one brain study and had to be upregulated in each study. Using the criteria, *TXLNG2P, USP9Y, UTY* and *CEBPA* were identified as plausible candidate genes (Table 3.9).

It should be noted that several limitations exist within the reviewed studies which have been considered. Many studies made use of already established datasets and only performed the subsequent differential gene expression analysis (Duke et al., 2007). This is problematic because if there is an error in the initial dataset, say with sample preparation or with the initial bioinformatics analysis, all subsequent results could include false positive findings. Study heterogeneity has been highlighted as a major problem (Dusonchet et al., 2014) as various platforms were utilized making it difficult for comparison between studies.

Therefore, it should be noted that the candidate list was selected based on the fact that the differentially expressed gene had to be present in both a blood and brain study thereby reducing the possibility of false positive findings and identification of genes that would be biologically relevant to PD pathobiology.

For gene expression studies it is assumed that the expression pattern for a gene can be used to predict the functional role of the encoded protein and therefore genes with similar expression profiles are thought to be related functionally. A major limitation highlighted by many studies and common to gene expression studies is to understand what the differential expression of certain genes means (Lockhart and Winzeler, 2000) . In other words, it is unknown whether a certain differentially expressed gene is directly involved in disease pathogenesis or whether in fact it is part of a downstream effect or process (Lu et al., 2006; Tan et al., 2005).

## 4.4 Biological pathways and enrichment analysis

The aim of the present study was to identify a ‘genetic signature’ or global cellular pathways which increased the risk of PD development in South African PD patients. To this end, IPA was used to assess the top canonical pathways and gene networks associated with the differentially expressed genes set generated. Regulation of eIF4 and p70S6K signaling was found to be the most significant deregulated canonical pathway found in whole blood of PD patients in this dataset (Table 3.7). The eIF4/p70S6K

pathway has been shown to be disrupted in PD neurons and is responsible for controlling energy metabolism in the cell (Elstner et al., 2011). Notably, the p70S6K signaling pathway suppresses autophagy allowing for the accumulation of dysfunctional proteins (Jiang et al., 2013) and *EIF4G1* mutations have also been identified in familial PD patients (Chartier-Harlin et al., 2011).

Interestingly both eIF4 and p70S6K signaling as well as EIF2 signaling are found to be deregulated throughout PD progression identified in the SN of PD patients with various Braak α-synuclein stages (range from 0-6) and could therefore aid in our understanding of the progression of this disease (Dijkstra et al., 2015).

The EIF2 signaling pathway was also identified in this dataset and in previous research (Elstner et al., 2011; Mutez et al., 2014; Scherzer et al., 2007b; Simunovic et al., 2010). These researchers identified the EIF2 pathway as one of the most dysregulated pathways in PBMCs and whole blood of sporadic and *LRRK2* mutation carriers. Mutez and colleagues allude to the fact that EIF2 is a new candidate pathway for PD (Mutez et al., 2014).

## 4.5 Link between *CEBPA*, *PGC1α* and PD

Of the 4 genes prioritized, *CEBPA* was selected for verification experiments. *CEBPA* encodes a protein which can modulate gene expression in cell cycle regulation. Mutations in *CEBPA* have been associated with acute myeloid leukemia (Libura et al., 2015; Sarojam et al., 2015) but little evidence exists for its involvement in PD besides being previously identified by two other genome-wide gene expression PD studies (Riley et al., 2014; Santiago and Potashkin, 2015).

*CEBPA* was highlighted in the gene expression network generated by IPA (Figure 3.9) which includes genes involved in cellular movement, cell to cell signaling and drug metabolism. In this figure, direct interaction between CEBPA, RNA polymerase II, CAM I, Histone H3 and Insulin are depicted by the solid lines. Of particular interest is the interaction between Insulin, CEBPA and peroxisome proliferator-activated receptor gamma coactivator-1α (*PGC-1α*). Notably, *PGC-1α* was also included in the filtered 132 candidate gene list (Table 3.5).

*PGC-1α*, as a master transcriptional regulator of mitochondrial biogenesis and oxidative metabolism, regulates transcription of a number of genes involved in mitochondrial protein import, mitochondrial protein folding and mitochondrial translation (Zheng et al., 2010). In a meta-analysis of genome wide expression studies performed by Zheng and colleagues, nine studies were included which had been performed on different brain tissue types, whole blood or lymphoblastoid cells (Zheng et al., 2010).

Most of these studies were included in our literature search and are presented in Table 1.3. Across both blood and brain studies, defects in the mitochondrial electron transport chain, glucose utilization and glucose sensing were found to occur early in PD pathogenesis.

Interestingly, in our study the most significant toxicity list generated through IPA was 'mitochondrial biogenesis'. The toxicology analysis is an added feature in IPA which allows for assessment of genes which may have a specific clinical phenotype that could potentially be toxic to the cell. It is of interest that mitochondrial biogenesis was highlighted and it could be speculated that the genes found to be dysregulated in PD patients in this study could affect the generation of new functional mitochondria.

It could further be speculated that there is a link between *CEBPA*, *PGC-1α* and PD, and that this could involve Insulin. Mounting evidence suggests that *CEBPA* could play a role in mitochondrial dysfunction. *CEBPA* has been shown to regulate gluconeogenesis (a biochemical process which results in glucose production) and lipogenesis (a process which involves the conversion of acetyl-CoA into fatty acids) (Matsusue et al., 2004). *CEBPA* was investigated in patients with Metabolic syndrome (MetS) and the study found that individuals with a specific genotype (A allele) had decreased insulin sensitivity and decreased insulin secretion (Delgado-Lista et al., 2013). Decreased insulin secretion and insulin sensitivity have been shown to cause impairment of peripheral glucose uptake thereby causing Type 2 diabetes (Guillausseau et al., 2008). It has been demonstrated that fatty acids cause insulin resistance. A study by Bonnard *et al* investigated the effect of a high fat, high sucrose diet (HFHSD) on mitochondrial density and functions, and demonstrated that a HFHSD caused mitochondrial alterations, increased ROS production and NADPH oxidase (Bonnard et al., 2008). Increased ROS production is of particular interest as this causes mitochondrial alterations in the muscle of diet-induced diabetic mice.

Also, mounting evidence suggests that mitochondrial dysfunction plays a key role in type 2 diabetes. Mitochondrial dysfunction is associated with increased insulin resistance and hyperglycemia which are prominent features of type 2 diabetes (Lowell and Shulman, 2005). And reduced mitochondrial oxidative phosphorylation has also been associated with Insulin resistance (Petersen et al., 2004). Changes in mitochondria morphology and reduced mitochondrial CI activity was observed in the skeletal muscle of type 2 diabetic patients (Kelley et al., 2002). Interestingly, reduced expression levels in a set of genes involved in oxidative phosphorylation was also observed in the muscle of type 2 diabetic and pre-diabetic subjects (Mootha et al., 2003; Patti et al., 2003)

Furthermore, previous studies have shown an association between type 2 diabetes and PD. It has been reported that around 50-80% of PD patients have abnormal glucose tolerance (Sandyk, 1993), which is a pre-diabetic state. Also, type 2 diabetes is associated with an increased risk of PD development

however, the mechanism behind this apparent association is unknown (Hu et al., 2007). It could be speculated that the *CEBPA* and *PGC-1α* genes could be implicated in this process.

As part of the larger parent study both patients and controls undergo extensive clinical examinations. Of interest in the larger study is the link between neuropsychiatric disorders and metabolic syndrome (MetS) therefore clinical information regarding whether study participants are diabetic or pre-diabetic is available. In total four PD patients (SR033, SR046, SR014 and SR010) and three controls (SR055, SR082 and SR111) were found to be pre-diabetic and three PD patients (SR005, SR023 and SR041) and one control (SR092) were diabetic according to the national guidelines for type 2 diabetes mellitus (Butler, 2009). Therefore, since the majority of patients and controls were not diabetic we do think that this variable had an influence on our study.

### *The proposed model of the link between CEBPA, PGC-1α, PD and Diabetes:*

From our findings, we hypothesize that there is a link between PD and diabetes through mitochondrial mechanisms which include mitochondrial dysfunction, mitochondrial biogenesis and mitochondrial oxidative phosphorylation (Figure 4.1). Due to various extrinsic and intrinsic factors such as ageing, sedentary lifestyles, diet, mitochondrial toxins and genetic mutations, there will be an accumulation of defective mitochondria in the cell. This will result in decreased ATP production with a concomitant increase in ROS levels. This in turn will cause oxidative stress that will ultimately lead to DA cell loss. (Indicated by the black arrows). Mitochondrial dysfunction also plays a role in causing hyperglycemia and increased insulin resistance which are key processes leading to type 2 diabetes. Increased insulin resistance and hyperglycemia also result in decreased mitochondrial oxidative phosphorylation which is linked to PD.

We speculate that in PD patients, the accumulation of unhealthy mitochondria is amplified, and this loss of healthy mitochondria causes upregulation of *PGC-1α* as a compensatory mechanism to increase mitochondrial biogenesis. However, new mitochondria are always produced from existing mitochondria (via the same 'pinching in half' process used by bacteria) and therefore if there is an accumulation of dysfunctional mitochondria, a large proportion of the new mitochondria will also be dysfunctional. This will lead to increased ROS and therefore increased oxidative stress causing DA neuronal loss. It is also plausible that *PGC-1α* regulates fatty acid oxidation and therefore causes increased fatty acid synthesis which impairs mitochondrial CI activity resulting in mitochondrial dysfunction.

Another possible mechanism of impaired CI activity and therefore mitochondrial dysfunction is through the increased production of fatty acids by lipogenesis triggered by *CEBPA*. Expression of *CEBPA* also causes decreased insulin sensitivity and secretion which causes increased Insulin resistance and ultimately diabetes. We do acknowledge however that this model is a highly simplistic and further work and validation studies are required to investigate other genes and the exact mechanisms underlying PD as well as the possible link between PD and diabetes.

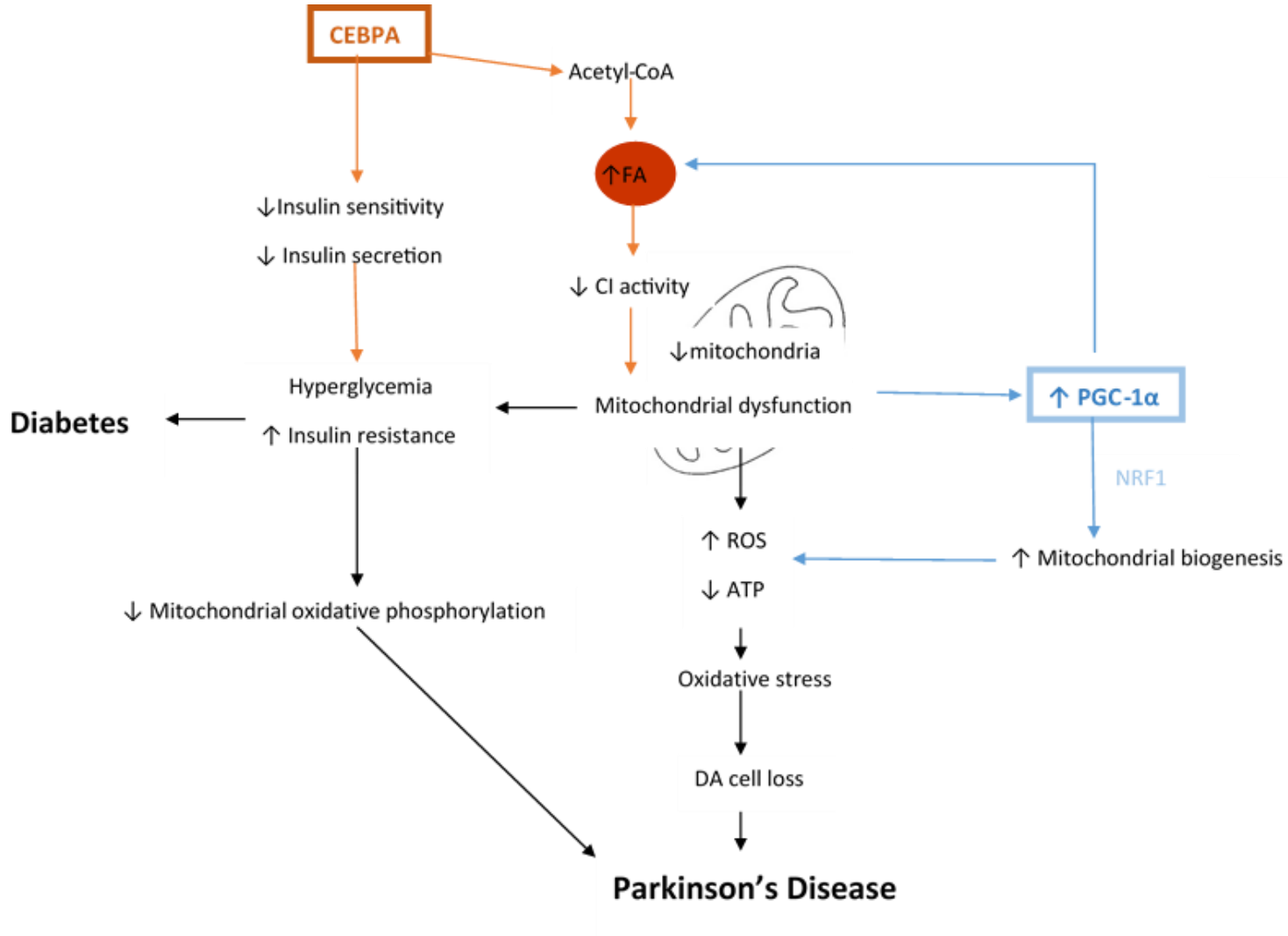



**Figure 4.1 Proposed model of the interaction between *CEBPA* and *PGC-1α* and the link between Parkinson's Disease and Diabetes mellitus.** Orange and blue arrows indicate *CEBPA* and *PGC-1α* mechanism of action respectively. Black arrows indicate the proposed mechanism involving mitochondrial dysfunction. ATP, adenosine triphosphate; *CEBPA*, CCAAT/enhancer binding protein alpha; CI, complex 1; DA, dopaminergic neurons; FA, fatty acids; *NRF1*, Nuclear respiratory factor 1; *PGC-1α,* Peroxisome proliferator-activated receptor gamma coactivator 1-alpha; ROS , reactive oxygen species.

## 4.6 Verification using qPCR

Data generated using qPCR on the Lightcycler® 96 software (Roche Diagnostics, Indianapolis, USA) was analysed using two different methods, REST-2009 © software and the $2^{-\Delta\Delta Cq}$ method, for comparative purposes. Overall, the findings from both methods correlated well.

*GAPDH* had a significant expression difference between patients and controls (Table 3.12) and it was therefore removed and the data re-analysed which demonstrated a significant down-regulation of the target gene (*CEBPA*) in patient samples compared to controls. Finding the appropriate reference gene is a major challenge for qPCR studies as highlighted by Coulson et al., 2008. Although selection of reference genes was chosen based on published literature from gene expression studies using blood and using the Ref gene network, this study has highlighted the fact that *GAPDH* may not be a suitable reference gene for qPCR studies done on whole blood. Also, *GAPDH* has been found to have variable expression in other studies and it is speculated that this could be due to the fact that it is involved in various cellular processes and could be influenced by different experimental conditions (Kozera and Rapacz, 2013). Therefore, it is important that reference genes are validated before use in a particular expression study.

*CEBPA* was shown to be significantly down-regulated in cases compared to controls and therefore this did not correlate with the RNA-Seq data in which up-regulation of *CEBPA* was observed in cases. Possible explanations for this discrepancy include the fact that the template cDNA concentration in the qPCR reactions were too low. Only 20 ng of cDNA template was used per reaction and the *CEBPA* Cq values obtained were between 35 and 41 cycles. These are higher than one would ideally expect for this type of experiment. Also, for some samples, the variation between triplicates was relatively high (Appendix XVII) due to problems with pipetting by the epMotion®5070f (Eppendorf, Hamburg, Germany) and this could skew the normalized expression values. The error bars reflected more variability between control samples compared to cases (Figure 3.13) this is due to the fact that sample preparation of controls was separate to that of the cases and therefore for future analyses sample preparation should be randomized to minimize variability between groups.

Furthermore, qPCR primers used for *CEBPA* in the present study may not be specific for the isoform detected in the RNA-Seq analysis. And finally, other possible reasons for the discrepancy in results is that the sample size is too small or *CEBPA* could be expressed at very low levels in blood, and therefore the findings are an artifact (false positive finding) in both experiments.

## 4.7 Study limitations

Limitations of the present study include the fact that only 40 study participants were utilized. However, compared to other published RNA-Seq studies on blood the number of study participants is reasonable. Other RNA-Seq ranged from a small sample size of 6 participants (3 PD and 3 controls) (Soreq et al., 2014) to larger groups of 60 individuals (Infante et al., 2016, 2015). The cases and controls were not gender matched in the present study as only female controls were available at the time of sample selection. This gender bias could skew the data. It should be noted that PD incidence is 1.5 times greater in males compared to females (Wooten et al., 2004) and therefore male patients are more likely to be recruited for research studies.

Another limitation of our study, and gene expression studies in general, is that mixed cell populations from blood were investigated. This brings into question whether studying blood is a good representative tissue for studying a neurological condition? The advantage of using blood is that it is easily obtainable from study participants and minimally invasive. Numerous studies have considered the importance of using whole blood to identify differential gene expression associated with the disease state and for identification of potential novel biomarkers in PD (Chatterjee et al., 2014; Chikina et al., 2015; Molochnikov et al., 2012; Santiago et al., 2013). Santiago and colleagues found that 7 out of 13 candidate biomarkers including hepatocyte nuclear factor (*HNF4A*) and polypyrimidine tract binding protein 1 (*PTBP1*) were dysregulated in whole blood of PD patients and are, therefore, potential PD biomarkers (Santiago et al., 2013). Whole blood mRNA signatures associated with the *LRRK2* mutant genotype in PD patients were also found (Chikina et al., 2015) as well as 23 novel miR markers (Chatterjee et al., 2014b).

There are a number of drawbacks with the use of brain tissue. Brain tissue is only accessible via autopsy postmortem and as this tissue is no longer viable the DA mechanism cannot be well explored (Grünblatt et al., 2004). Also, in PD patients 60-80% of the DA neurons in the SN have died and therefore, the findings on post-mortem SN tissue may actually be representative of surviving potentially healthy DA neurons and not the abnormalities that arise in neurons which are lost in PD (Cantuti-Castelvetri et al., 2007). Another limitation is that isolation of the specific region of interest is complicated and therefore the differentially expressed genes could rather be representative of the surrounding glial cells and not only of DA neurons (Moran et al., 2006).

Moreover, it has been shown that whole blood and brain tissue share significant gene expression similarities (Liew et al., 2006; Sullivan et al., 2006). Therefore, regulatory mechanisms in these tissues could be very similar and gene expression signals could be used in the development of blood biomarkers for PD (Scherzer et al., 2007). RNA-Seq as a gene expression profiling tool offers a potential avenue in identifying blood-based biomarkers for PD.

Another limitation of this study which may not be that obvious is that the reads were mapped to the reference hg19. The mixed ancestry population is of a unique admixture of diverse populations (de Wit et al., 2010) and therefore by mapping reads to a European reference, unique transcripts and isoforms could potentially be missed. There is therefore an urgent need for a South African mixed ancestry reference genome. Re-analysis of data with mapping to this reference genome could highlight genes involved in PD pathogenesis which have been missed using the generic reference.

Also, bioinformatic tools are constantly being updated and new ones are being developed and therefore selection of the 'best' pipeline proves challenging. In this study although STAR was shown to be a good alignment tool other tools exist such as GSNAP, GSTRUCT and MapSplice compare favorably(Engström et al., 2013; Grant et al., 2011). A recent article highlighted the use of DE-Seq for quantification (Soneson and Delorenzi, 2013) and should the RNA-Seq data be re-analysed I would recommend replacing GSA in our pipeline. Bioinformatic analysis can only be performed with the tools available at the time of which the researcher deems most appropriate however, study comparisons prove challenging as different researchers use different tools based on different algorithms.

Previously highlighted limitations of RNA-Seq include the bioinformatic challenges and the challenge to store, retrieve and process 'big' data. RNA-Seq generally produces short reads and mapping these reads to a reference proves challenging as reads could span exon junctions or contain poly (A) tails which may not have been removed effectively. Reads could also match multiple locations in the genome and perhaps by using sequencing platforms such as the Solid 454 which produces longer reads may help to locate the exact position of these reads. High sequence coverage is also required to detect rare transcripts but the higher the coverage, the higher the cost.

## 4.8 Future Work

Once raw RNA-Seq reads have been generated multiple analyses can be performed to answer specific biological questions. Therefore, re-analysis of raw data could be performed using Linnux allowing more control over individual analysis steps to investigate events such as alternative splicing. Also, specific RNA cell populations can be investigated including miRNA and lncRNAs.

At the time of sample selection only 40 study participants were available and only female controls had been recruited. Future work should include a larger sample size (a minimum of 100 study participants, 50 cases and 50 controls) that are age and sex matched. It is recommended that sample matching (age, sex, ethnicity and socioeconomic status using years of schooling as a proxy) of patients and controls should be in the form of pair matching, where each patient is matched individually to a control, rather than taking the average over the entire group. Also, in addition to MLPA the samples should be screened for pathogenic mutations in known PD genes using a targeted resequencing panel such as the Ion AmpliSeq™ Inherited Disease Panel (ThermoFisher Scientific, MA, USA). This will enable the patients to be sub-divided into genetic and idiopathic forms of PD prior to RNA-Seq analysis.

As part of the larger parent study Shared Roots, hair and dermal fibroblast samples have been obtained from all study participants. Conversion of fibroblasts to iPSCs could allow functional studies on neuronal cell lines derived from patients in order to further investigate possible disease pathways identified using RNA-Seq analysis. Integration of genomic, transcriptomic and epigenetic data would facilitate insight into a complex disease such as PD. The parent study also involves the recruitment of cases and controls with and without metabolic syndrome (MetS). Based on the results obtained from this pilot study (which include a potential mechanism of the *CEBPA* gene and the link to diabetes), it would be interesting to perform RNA-Seq on PD patients and controls with and without MetS and to compare differentially expressed genes and pathways between the two groups. Investigation between these two groups could provide a better understanding of the link between diseases associated with lifestyle changes and PD.

Also, the RNA-Seq technique is constantly being further developed and improved upon and one such exciting development is called single cell RNA-Seq (reviewed by Wu et al., 2014). This technique involves the dissociation of solid tissue into single cells which are isolated by laser-capture microdissection. RNA is then isolated and converted to cDNA or amplified and then converted. Adaptors are then ligated to one or both ends of the molecule, sequenced and mapped to the reference genome. Clustering of expression profiles allows for identification of specific cell types which can be traced back to a specific location in the tissue. This technique therefore allows quantification of the transcriptome of individual cells with a small number of sequencing reads (Wu et al., 2014). Another new technique recently published by Stahl and colleagues involves the detection of transcripts in situ by using laser capture to extract small tissue patches and barcode each sample individually (Ståhl et al., 2016). Reads can therefore be linked to specific locations in the cell. This means that in the future transcriptional profiling can be done on individual cells of interest or even within selected regions within individual cells.

## 4.9 Conclusions

Transcriptomic profiling using RNA-Seq revealed differential gene expression between PD patients and controls. To our knowledge, this the first RNA-Seq study performed on South African PD patients and only the sixth RNA-Seq study performed on PD patients worldwide. The five published studies are as follows: Gollamudi et al., 2012; Hossein-Nezhad et al., 2016; Infante et al., 2016, 2015; Soreq et al., 2014. Other strengths of our study include investigation of PD in a very unique population and generation of good quality raw RNA-Seq data that can be re-analyzed as bioinformatics pipelines are improved. We recommend that prior to RNA-seq, initial mutation screening of all PD patients should be performed to identify patients with pathogenic mutations as they may have a different gene expression profile to that of idiopathic PD.

Based on the results obtained, RNA-Seq is a powerful technique for identification of novel disease pathways in PD. Although the sample size was small in the present study and sex biased, candidate genes were identified that warrant further investigation. The interesting finding of a possible link between diabetes-associated genes *CEBPA* and *PGC-1α* and PD is exciting, and not implausible as patients with type 2 diabetes have been shown to be at increased risk for PD development (Hu et al., 2007). If validated in follow-up studies, it is possible that these transcripts and proteins may be potential blood biomarkers for early detection of individuals at risk of development of PD, for monitoring of disease progression or for monitoring of pharmacologic responses to a therapeutic intervention.

# Reference List

## Articles:

Alam, M.A., Rahman, M.M., 2014. Mitochondrial dysfunction in obesity: potential benefit and mechanism of Co-enzyme Q10 supplementation in metabolic syndrome. J. Diabetes Metab. Disord. 13, 60. doi:10.1186/2251-6581-13-60

Alieva, A.K., Shadrina, M.I., Filatova, E.V., Karabanov, A.V., Illarioshkin, S.N., Limborska, S.A., Slominsky, P.A., 2014. Involvement of endocytosis and alternative splicing in the formation of the pathological process in the early stages of Parkinson's disease. BioMed Res. Int. 2014, 718732. doi:10.1155/2014/718732

Bach, J.-P., Ziegler, U., Deuschl, G., Dodel, R., Doblhammer-Reiter, G., 2011. Projected numbers of people with movement disorders in the years 2030 and 2050. Mov. Disord. Off. J. Mov. Disord. Soc. 26, 2286–2290. doi:10.1002/mds.23878

Bardien, S., Keyser, R., Yako, Y., Lombard, D., Carr, J., 2009. Molecular analysis of the parkin gene in South African patients diagnosed with Parkinson's disease. Parkinsonism Relat. Disord. 15, 116–121. doi:10.1016/j.parkreldis.2008.04.005

Basso, M., Giraudo, S., Corpillo, D., Bergamasco, B., Lopiano, L., Fasano, M., 2004. Proteome analysis of human substantia nigra in Parkinson's disease. Proteomics 4, 3943–3952. doi:10.1002/pmic.200400848

Betarbet, R., Sherer, T.B., MacKenzie, G., Garcia-Osuna, M., Panov, A.V., Greenamyre, J.T., 2000. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. Nat. Neurosci. 3, 1301–1306. doi:10.1038/81834

Bezard, E., Imbert, C., Deloire, X., Bioulac, B., Gross, C.E., 1997. A chronic MPTP model reproducing the slow evolution of Parkinson's disease: evolution of motor symptoms in the monkey. Brain Res. 766, 107–112. doi:10.1016/S0006-8993(97)00531-3

Bindoff, L.A., Birch-Machin, M.A., Cartlidge, N.E., Parker, W.D., Turnbull, D.M., 1991. Respiratory chain abnormalities in skeletal muscle from patients with Parkinson's disease. J. Neurol. Sci. 104, 203–208.

Biskup, S., Moore, D.J., Celsi, F., Higashi, S., West, A.B., Andrabi, S.A., Kurkinen, K., Yu, S.-W., Savitt, J.M., Waldvogel, H.J., Faull, R.L.M., Emson, P.C., Torp, R., Ottersen, O.P., Dawson, T.M., Dawson, V.L., 2006. Localization of LRRK2 to membranous and vesicular structures in mammalian brain. Ann. Neurol. 60, 557–569. doi:10.1002/ana.21019

Blesa, J., Trigo-Damas, I., Quiroga-Varela, A., Jackson-Lewis, V.R., 2015. Oxidative stress and Parkinson's disease. Front. Neuroanat. 9. doi:10.3389/fnana.2015.00091

Bonifacino, J.S., Rojas, R., 2006. Retrograde transport from endosomes to the trans-Golgi network. Nat. Rev. Mol. Cell Biol. 7, 568–579. doi:10.1038/nrm1985

Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C.J., Squitieri, F., Ibanez, P., Joosse, M., van Dongen, J.W., Vanacore, N., van Swieten, J.C., Brice, A., Meco, G., van Duijn, C.M., Oostra, B.A., Heutink, P., 2003. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299, 256–259. doi:10.1126/science.1077209

Bonnard, C., Durand, A., Peyrol, S., Chanseaume, E., Chauvin, M.-A., Morio, B., Vidal, H., Rieusset, J., 2008. Mitochondrial dysfunction results from oxidative stress in the skeletal muscle of diet-induced insulin-resistant mice. J. Clin. Invest. 118, 789–800. doi:10.1172/JCI32601

Bossers, K., Meerhoff, G., Balesar, R., van Dongen, J.W., Kruse, C.G., Swaab, D.F., Verhaagen, J., 2009a. Analysis of gene expression in Parkinson's disease: possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol. Zurich Switz. 19, 91–107. doi:10.1111/j.1750-3639.2008.00171.x

Bossers, K., Meerhoff, G., Balesar, R., van Dongen, J.W., Kruse, C.G., Swaab, D.F., Verhaagen, J., 2009b. Analysis of gene expression in Parkinson's disease: possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol. Zurich Switz. 19, 91–107. doi:10.1111/j.1750-3639.2008.00171.x

Botta-Orfila, T., Morató, X., Compta, Y., Lozano, J.J., Falgàs, N., Valldeoriola, F., Pont-Sunyer, C., Vilas, D., Mengual, L., Fernández, M., Molinuevo, J.L., Antonell, A., Martí, M.J., Fernández-Santiago, R., Ezquerra, M., 2014. Identification of blood serum micro-RNAs associated with idiopathic and LRRK2 Parkinson's disease. J. Neurosci. Res. 92, 1071–1077. doi:10.1002/jnr.23377

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., Tredici, K.D., 2004. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res. 318, 121–134. doi:10.1007/s00441-004-0956-9

Braak, H., Tredici, K.D., Rüb, U., de Vos, R.A.I., Jansen Steur, E.N.H., Braak, E., 2003. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol. Aging 24, 197–211. doi:10.1016/S0197-4580(02)00065-9

Brüggemann, N., Mitterer, M., Lanthaler, A.J., Djarmati, A., Hagenah, J., Wiegers, K., Winkler, S., Pawlack, H., Lohnau, T., Pramstaller, P.P., Klein, C., Lohmann, K., 2009. Frequency of heterozygous Parkin mutations in healthy subjects: need for careful prospective follow-up examination of mutation carriers. Parkinsonism Relat. Disord. 15, 425–429. doi:10.1016/j.parkreldis.2008.11.014

Burchell, V.S., Nelson, D.E., Sanchez-Martinez, A., Delgado-Camprubi, M., Ivatt, R.M., Pogson, J.H., Randle, S.J., Wray, S., Lewis, P.A., Houlden, H., Abramov, A.Y., Hardy, J., Wood, N.W., Whitworth, A.J., Laman, H., Plun-Favreau, H., 2013. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. Nat. Neurosci. 16, 1257–1265. doi:10.1038/nn.3489

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622. doi:10.1373/clinchem.2008.112797

Butler, N., 2009. National Guidelines at a glance: Type 2 Diabetes Mellitus. SA Pharm. J. 76, 32–35, 43.

Canet-Avilés, R.M., Wilson, M.A., Miller, D.W., Ahmad, R., McLendon, C., Bandyopadhyay, S., Baptista, M.J., Ringe, D., Petsko, G.A., Cookson, M.R., 2004. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proc. Natl. Acad. Sci. U. S. A. 101, 9103–9108. doi:10.1073/pnas.0402959101

Cantuti-Castelvetri, I., Keller-McGandy, C., Bouzou, B., Asteris, G., Clark, T.W., Frosch, M.P., Standaert, D.G., 2007. Effects of gender on nigral gene expression and parkinson disease. Neurobiol. Dis. 26, 606–614. doi:10.1016/j.nbd.2007.02.009

Chan, N.C., Chan, D.C., 2011. Parkin uses the UPS to ship off dysfunctional mitochondria. Autophagy 7, 771–772. doi:10.4161/auto.7.7.15453

Chance, B., Sies, H., Boveris, A., 1979. Hydroperoxide metabolism in mammalian organs. Physiol. Rev. 59, 527–605.

Chang, Y.-F., Cheng, C.-M., Chang, L.-K., Jong, Y.-J., Yuo, C.-Y., 2006. The F-box protein Fbxo7 interacts with human inhibitor of apoptosis protein cIAP1 and promotes cIAP1 ubiquitination. Biochem. Biophys. Res. Commun. 342, 1022–1026. doi:10.1016/j.bbrc.2006.02.061

Chartier-Harlin, M.-C., Dachsel, J.C., Vilariño-Güell, C., Lincoln, S.J., Leprêtre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.-S., Le Rhun, E., Mutez, E., Larvor, L., Duflot, A., Vanbesien-Mailliot, C., Kreisler, A., Ross, O.A., Nishioka, K., Soto-Ortolaza, A.I., Cobb, S.A., Melrose, H.L., Behrouz, B., Keeling, B.H., Bacon, J.A., Hentati, E., Williams, L., Yanagiya, A., Sonenberg, N., Lockhart, P.J., Zubair, A.C., Uitti, R.J., Aasly, J.O., Krygowska-Wajs, A., Opala, G., Wszolek, Z.K., Frigerio, R., Maraganore, D.M., Gosal, D., Lynch, T., Hutchinson, M., Bentivoglio, A.R., Valente, E.M., Nichols, W.C., Pankratz, N., Foroud, T., Gibson, R.A., Hentati, F., Dickson,

D.W., Destée, A., Farrer, M.J., 2011. Translation Initiator EIF4G1 Mutations in Familial Parkinson Disease. Am. J. Hum. Genet. 89, 398–406. doi:10.1016/j.ajhg.2011.08.009

Chatterjee, P., Bhattacharyya, M., Bandyopadhyay, S., Roy, D., 2014a. Studying the system-level involvement of microRNAs in Parkinson's disease. PloS One 9, e93751. doi:10.1371/journal.pone.0093751

Chatterjee, P., Bhattacharyya, M., Bandyopadhyay, S., Roy, D., 2014b. Studying the System-Level Involvement of MicroRNAs in Parkinson's Disease. PLoS ONE 9, e93751. doi:10.1371/journal.pone.0093751

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., Varshavsky, A., 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243, 1576–1583.

Chaudhuri, K.R., Martinez-Martin, P., Brown, R.G., Sethi, K., Stocchi, F., Odin, P., Ondo, W., Abe, K., Macphee, G., Macmahon, D., Barone, P., Rabey, M., Forbes, A., Breen, K., Tluk, S., Naidu, Y., Olanow, W., Williams, A.J., Thomas, S., Rye, D., Tsuboi, Y., Hand, A., Schapira, A.H.V., 2007. The metric properties of a novel non-motor symptoms scale for Parkinson's disease: Results from an international pilot study. Mov. Disord. Off. J. Mov. Disord. Soc. 22, 1901–1911. doi:10.1002/mds.21596

Chikina, M.D., Gerald, C.P., Li, X., Ge, Y., Pincas, H., Nair, V.D., Wong, A.K., Krishnan, A., Troyanskaya, O.G., Raymond, D., Saunders-Pullman, R., Bressman, S.B., Yue, Z., Sealfon, S.C., 2015. Low-variance RNAs identify Parkinson's disease molecular signature in blood. Mov. Disord. Off. J. Mov. Disord. Soc. 30, 813–821. doi:10.1002/mds.26205

Choi, P., Snyder, H., Petrucelli, L., Theisler, C., Chong, M., Zhang, Y., Lim, K., Chung, K.K.K., Kehoe, K., D'Adamio, L., Lee, J.M., Cochran, E., Bowser, R., Dawson, T.M., Wolozin, B., 2003. SEPT5_v2 is a parkin-binding protein. Brain Res. Mol. Brain Res. 117, 179–189.

Chu, Y., Corey, D.R., 2012. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. Nucleic Acid Ther. 22, 271–274. doi:10.1089/nat.2012.0367

Chung, K.K.K., Zhang, Y., Lim, K.L., Tanaka, Y., Huang, H., Gao, J., Ross, C.A., Dawson, V.L., Dawson, T.M., 2001. Parkin ubiquitinates the α-synuclein–interacting protein, synphilin-1: implications for Lewy-body formation in Parkinson disease. Nat. Med. 7, 1144–1150. doi:10.1038/nm1001-1144

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K., Lander, E.S., 2007. Distinguishing protein-coding and noncoding genes in the human genome. Proc. Natl. Acad. Sci. 104, 19428–19433. doi:10.1073/pnas.0709013104

Cochemé, H.M., Murphy, M.P., 2008. Complex I is the major site of mitochondrial superoxide production by paraquat. J. Biol. Chem. 283, 1786–1798. doi:10.1074/jbc.M708597200

Corradini, B.R., Iamashita, P., Tampellini, E., Farfel, J., Marcelo, Grinberg, L.T., Moreira-Filho, C.A., Corradini, B.R., Iamashita, P., Tampellini, E., Farfel, J., Marcelo, Grinberg, L.T., Moreira-Filho, C.A., 2014. Complex Network-Driven View of Genomic Mechanisms Underlying Parkinson’s Disease: Analyses in Dorsal Motor Vagal Nucleus, Locus Coeruleus, and Substantia Nigra, Complex Network-Driven View of Genomic Mechanisms Underlying Parkinson’s Disease: Analyses in Dorsal Motor Vagal Nucleus, Locus Coeruleus, and Substantia Nigra. BioMed Res. Int. BioMed Res. Int. 2014, 2014, e543673. doi:10.1155/2014/543673, 10.1155/2014/543673

Coulson, D.T., Brockbank, S., Quinn, J.G., Murphy, S., Ravid, R., Irvine, G.B., Johnston, J.A., 2008. Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. BMC Mol. Biol. 9, 46. doi:10.1186/1471-2199-9-46

Dagda, R.K., Cherra, S.J., Kulich, S.M., Tandon, A., Park, D., Chu, C.T., 2009. Loss of PINK1 function promotes mitophagy through effects on oxidative stress and mitochondrial fission. J. Biol. Chem. 284, 13843–13855. doi:10.1074/jbc.M808515200

Dauer, W., Przedborski, S., 2003. Parkinson's Disease: Mechanisms and Models. Neuron 39, 889–909. doi:10.1016/S0896-6273(03)00568-3

Davis, G.C., Williams, A.C., Markey, S.P., Ebert, M.H., Caine, E.D., Reichert, C.M., Kopin, I.J., 1979. Chronic Parkinsonism secondary to intravenous injection of meperidine analogues. Psychiatry Res. 1, 249–254.

de Lau, L.M., Breteler, M.M., 2006. Epidemiology of Parkinson's disease. Lancet Neurol. 5, 525–535. doi:10.1016/S1474-4422(06)70471-9

de Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P.D., Seoighe, C., Hoal, E.G., 2010. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum. Genet. 128, 145–153. doi:10.1007/s00439-010-0836-1

Delgado-Lista, J., Perez-Martinez, P., Garcia-Rios, A., Phillips, C.M., Hall, W., Gjelstad, I.M.F., Lairon, D., Saris, W., Kieć-Wilk, B., Karlström, B., Drevon, C.A., Defoort, C., Blaak, E.E., Dembinska-Kieć, A., Risérus, U., Lovegrove, J.A., Roche, H.M., Lopez-Miranda, J., 2013. A gene variation (rs12691) in the CCAT/enhancer binding protein α modulates glucose metabolism in metabolic syndrome. Nutr. Metab. Cardiovasc. Dis. NMCD 23, 417–423. doi:10.1016/j.numecd.2011.09.008

Deng, H.-X., Shi, Y., Yang, Y., Ahmeti, K.B., Miller, N., Huang, C., Cheng, L., Zhai, H., Deng, S., Nuytemans, K., Corbett, N.J., Kim, M.J., Deng, H., Tang, B., Yang, Z., Xu, Y., Chan, P., Huang, B., Gao, X.-P., Song, Z., Liu, Z., Fecto, F., Siddique, N., Foroud, T., Jankovic, J., Ghetti, B., Nicholson, D.A., Krainc, D., Melen, O., Vance, J.M., Pericak-Vance, M.A., Ma, Y.-C., Rajput, A.H., Siddique, T., 2016. Identification of TMEM230 mutations in familial Parkinson's disease. Nat. Genet. 48, 733–739. doi:10.1038/ng.3589

Desai, V.G., Feuers, R.J., Hart, R.W., Ali, S.F., 1996. MPP(+)-induced neurotoxicity in mouse is age-dependent: evidenced by the selective inhibition of complexes of electron transport. Brain Res. 715, 1–8. doi:10.1016/0006-8993(95)01255-9

Devi, L., Raghavendran, V., Prabhu, B.M., Avadhani, N.G., Anandatheerthavarada, H.K., 2008. Mitochondrial import and accumulation of alpha-synuclein impair complex I in human dopaminergic neuronal cultures and Parkinson disease brain. J. Biol. Chem. 283, 9089–9100. doi:10.1074/jbc.M710012200

Dexter, D.T., Carter, C.J., Wells, F.R., Javoy-Agid, F., Agid, Y., Lees, A., Jenner, P., Marsden, C.D., 1989. Basal lipid peroxidation in substantia nigra is increased in Parkinson's disease. J. Neurochem. 52, 381–389.

Dias, V., Junn, E., Mouradian, M.M., 2013. The Role of Oxidative Stress in Parkinson's Disease. J. Park. Dis. 3, 461–491. doi:10.3233/JPD-130230

Dickson, D.W., 2012. Parkinson's Disease and Parkinsonism: Neuropathology. Cold Spring Harb. Perspect. Med. 2. doi:10.1101/cshperspect.a009258

Dickson, D.W., 2007. Linking Selective Vulnerability to Cell Death Mechanisms in Parkinson's Disease. Am. J. Pathol. 170, 16–19. doi:10.2353/ajpath.2007.061011

Dijkstra, A.A., Ingrassia, A., de Menezes, R.X., van Kesteren, R.E., Rozemuller, A.J.M., Heutink, P., van de Berg, W.D.J., 2015. Evidence for Immune Response, Axonal Dysfunction and Reduced Endocytosis in the Substantia Nigra in Early Stage Parkinson's Disease. PloS One 10, e0128651. doi:10.1371/journal.pone.0128651

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. 29, 15–21. doi:10.1093/bioinformatics/bts635

Dong, Z., Ferger, B., Paterna, J.-C., Vogel, D., Furler, S., Osinde, M., Feldon, J., Büeler, H., 2003. Dopamine-dependent neurodegeneration in rats induced by viral vector-mediated overexpression of the parkin target protein, CDCrel-1. Proc. Natl. Acad. Sci. U. S. A. 100, 12438–12443. doi:10.1073/pnas.2132992100

Dotchin, C., Jusabani, A., Gray, W.K., Walker, R., 2012. Projected numbers of people with movement disorders in the years 2030 and 2050: Implications for sub-Saharan Africa, using essential tremor and Parkinson's disease in Tanzania as an example. Mov. Disord. 27, 1204–1205. doi:10.1002/mds.25097

Doty, R.L., 2012. Olfactory dysfunction in Parkinson disease. Nat. Rev. Neurol. 8, 329–339. doi:10.1038/nrneurol.2012.80

Drouet, V., rie, Lesage, S., Drouet, V., rie, Lesage, S., 2014. Synaptojanin 1 Mutation in Parkinson’s Disease Brings Further Insight into the Neuropathological Mechanisms, Synaptojanin 1 Mutation in Parkinson’s Disease Brings Further Insight into the Neuropathological Mechanisms. BioMed Res. Int. BioMed Res. Int. 2014, 2014, e289728. doi:10.1155/2014/289728, 10.1155/2014/289728

Duke, D.C., Moran, L.B., Pearce, R.K.B., Graeber, M.B., 2007. The medial and lateral substantia nigra in Parkinson's disease: mRNA profiles associated with higher brain tissue vulnerability. Neurogenetics 8, 83–94. doi:10.1007/s10048-006-0077-6

Dusonchet, J., Li, H., Guillily, M., Liu, M., Stafa, K., Derada Troletti, C., Boon, J.Y., Saha, S., Glauser, L., Mamais, A., Citro, A., Youmans, K.L., Liu, L., Schneider, B.L., Aebischer, P., Yue, Z., Bandopadhyay, R., Glicksman, M.A., Moore, D.J., Collins, J.J., Wolozin, B., 2014. A Parkinson's disease gene regulatory network identifies the signaling protein RGS2 as a modulator of LRRK2 activity and neuronal toxicity. Hum. Mol. Genet. 23, 4887–4905. doi:10.1093/hmg/ddu202

Dutta, A., Le Magnen, C., Mitrofanova, A., Ouyang, X., Califano, A., Abate-Shen, C., 2016. Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. Science 352, 1576–1580. doi:10.1126/science.aad9512

Eeden, S.K.V.D., Tanner, C.M., Bernstein, A.L., Fross, R.D., Leimpeter, A., Bloch, D.A., Nelson, L.M., 2003. Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. Am. J. Epidemiol. 157, 1015–1022. doi:10.1093/aje/kwg068

Elstner, M., Morris, C.M., Heim, K., Bender, A., Mehta, D., Jaros, E., Klopstock, T., Meitinger, T., Turnbull, D.M., Prokisch, H., 2011. Expression analysis of dopaminergic neurons in Parkinson's disease and aging links transcriptional dysregulation of energy metabolism to cell death. Acta Neuropathol. (Berl.) 122, 75–86. doi:10.1007/s00401-011-0828-9

Engström, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Rätsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigó, R., Bertone, P., RGASP Consortium, 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10, 1185–1191. doi:10.1038/nmeth.2722

Fabre, E., Monserrat, J., Herrero, A., Barja, G., Leret, M.L., 1999. Effect of MPTP on brain mitochondrial H2O2 and ATP production and on dopamine and DOPAC in the striatum. J. Physiol. Biochem. 55, 325–331.

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., Tomiyama, H., Nishioka, K., Hasegawa, K., Saiki, H., Satake, W., Mogushi, K., Sasaki, R., Kokubo, Y., Kuzuhara, S., Toda, T., Mizuno, Y., Uchiyama, Y., Ohno, K., Hattori, N., 2015. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. Lancet Neurol. 14, 274–282. doi:10.1016/S1474-4422(14)70266-2

Gandhi, S., Wood-Kaczmar, A., Yao, Z., Plun-Favreau, H., Deas, E., Klupsch, K., Downward, J., Latchman, D.S., Tabrizi, S.J., Wood, N.W., Duchen, M.R., Abramov, A.Y., 2009. PINK1-associated Parkinson's disease is caused by neuronal vulnerability to calcium-induced cell death. Mol. Cell 33, 627–638. doi:10.1016/j.molcel.2009.02.013

Gan-Or, Z., Dion, P.A., Rouleau, G.A., 2015. Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. Autophagy 11, 1443–1457. doi:10.1080/15548627.2015.1067364

Gardet, A., Benita, Y., Li, C., Sands, B.E., Ballester, I., Stevens, C., Korzenik, J.R., Rioux, J.D., Daly, M.J., Xavier, R.J., Podolsky, D.K., 2010. LRRK2 is involved in the IFN-gamma response and host response to pathogens. J. Immunol. Baltim. Md 1950 185, 5577–5585. doi:10.4049/jimmunol.1000548

Geisler, S., Holmström, K.M., Skujat, D., Fiesel, F.C., Rothfuss, O.C., Kahle, P.J., Springer, W., 2010. PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. Nat. Cell Biol. 12, 119–131. doi:10.1038/ncb2012

Gollamudi, S., Johri, A., Calingasan, N.Y., Yang, L., Elemento, O., Beal, M.F., 2012. Concordant signaling pathways produced by pesticide exposure in mice correspond to pathways identified in human Parkinson's disease. PloS One 7, e36191. doi:10.1371/journal.pone.0036191

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., Pierce, E.A., 2011. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinforma. Oxf. Engl. 27, 2518–2528. doi:10.1093/bioinformatics/btr427

Greenamyre, J.T., MacKenzie, G., Peng, T.I., Stephans, S.E., 1999a. Mitochondrial dysfunction in Parkinson's disease. Biochem. Soc. Symp. 66, 85–97.

Greenamyre, J.T., MacKenzie, G., Peng, T.I., Stephans, S.E., 1999b. Mitochondrial dysfunction in Parkinson's disease. Biochem. Soc. Symp. 66, 85–97.

Grünblatt, E., Mandel, S., Jacob-Hirsch, J., Zeligson, S., Amariglo, N., Rechavi, G., Li, J., Ravid, R., Roggendorf, W., Riederer, P., Youdim, M.B.H., 2004. Gene expression profiling of parkinsonian substantia nigra pars compacta; alterations in ubiquitin-proteasome, heat shock protein, iron and oxidative stress regulated proteins, cell adhesion/cellular matrix and vesicle trafficking genes. J. Neural Transm. Vienna Austria 1996 111, 1543–1573. doi:10.1007/s00702-004-0212-1

Grünblatt, E., Zander, N., Bartl, J., Jie, L., Monoranu, C.-M., Arzberger, T., Ravid, R., Roggendorf, W., Gerlach, M., Riederer, P., 2007. Comparison analysis of gene expression patterns between sporadic Alzheimer's and Parkinson's disease. J. Alzheimers Dis. JAD 12, 291–311.

Guillausseau, P.-J., Meas, T., Virally, M., Laloi-Michelin, M., Médeau, V., Kevorkian, J.-P., 2008. Abnormalities in insulin secretion in type 2 diabetes mellitus. Diabetes Metab. 34 Suppl 2, S43-48. doi:10.1016/S1262-3636(08)73394-9

Haas, R.H., Nasirian, F., Nakano, K., Ward, D., Pay, M., Hill, R., Shults, C.W., 1995. Low platelet mitochondrial complex I and complex II/III activity in early untreated Parkinson's disease. Ann. Neurol. 37, 714–722. doi:10.1002/ana.410370604

Hamilton, M., 1960. A RATING SCALE FOR DEPRESSION. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

Hauser, M.A., Li, Y.-J., Takeuchi, S., Walters, R., Noureddine, M., Maready, M., Darden, T., Hulette, C., Martin, E., Hauser, E., Xu, H., Schmechel, D., Stenger, J.E., Dietrich, F., Vance, J., 2003. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. Hum. Mol. Genet. 12, 671–677.

Hauser, M.A., Li, Y.-J., Xu, H., Noureddine, M.A., Shao, Y.S., Gullans, S.R., Scherzer, C.R., Jensen, R.V., McLaurin, A.C., Gibson, J.R., Scott, B.L., Jewett, R.M., Stenger, J.E., Schmechel, D.E., Hulette, C.M., Vance, J.M., 2005. Expression profiling of substantia nigra in Parkinson disease, progressive supranuclear palsy, and frontotemporal dementia with parkinsonism. Arch. Neurol. 62, 917–921. doi:10.1001/archneur.62.6.917

Hawkes, C.H., Shephard, B.C., Daniel, S.E., 1999. Is Parkinson's disease a primary olfactory disorder? QJM 92, 473–480. doi:10.1093/qjmed/92.8.473

Haylett, W.L., Keyser, R.J., Plessis, M.C. du, Merwe, C. van der, Blanckenberg, J., Lombard, D., Carr, J., Bardien, S., 2012. Mutations in the parkin gene are a minor cause of Parkinson's disease in the South African population. Parkinsonism Relat. Disord. 18, 89–92. doi:10.1016/j.parkreldis.2011.09.022

Hershko, A., Ciechanover, A., 1998. The ubiquitin system. Annu. Rev. Biochem. 67, 425–479. doi:10.1146/annurev.biochem.67.1.425

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., Pringsheim, T., 2016. The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. Neuroepidemiology 46, 292–300. doi:10.1159/000445751

Hoppins, S., Lackner, L., Nunnari, J., 2007. The machines that divide and fuse mitochondria. Annu. Rev. Biochem. 76, 751–780. doi:10.1146/annurev.biochem.76.071905.090048

Hossein-Nezhad, A., Fatemi, R.P., Ahmad, R., Peskind, E.R., Zabetian, C.P., Hu, S.-C., Shi, M., Wahlestedt, C., Zhang, J., Faghihi, M.A., 2016. Transcriptomic Profiling of Extracellular RNAs Present in

Cerebrospinal Fluid Identifies Differentially Expressed Transcripts in Parkinson's Disease. J. Park. Dis. 6, 109–117. doi:10.3233/JPD-150737

Hruz, T., Wyss, M., Docquier, M., Pfaffl, M.W., Masanetz, S., Borghi, L., Verbrugghe, P., Kalaydjieva, L., Bleuler, S., Laule, O., Descombes, P., Gruissem, W., Zimmermann, P., 2011. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 12, 156. doi:10.1186/1471-2164-12-156

Hu, G., Jousilahti, P., Bidel, S., Antikainen, R., Tuomilehto, J., 2007. Type 2 Diabetes and the Risk of Parkinson's Disease. Diabetes Care 30, 842–847. doi:10.2337/dc06-2011

Hu, M., Cooper, J., Beamish, R., Jones, E., Butterworth, R., Catterall, L., Ben-shlomo, Y., 2011. How well do we recognise non-motor symptoms in a British Parkinson's disease population? J. Neurol. 258, 1513–7. doi:http://dx.doi.org/10.1007/s00415-011-5972-6

Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J., 1992. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry 55, 181–184. doi:10.1136/jnnp.55.3.181

Hwang, O., 2013. Role of oxidative stress in Parkinson's disease. Exp. Neurobiol. 22, 11–17. doi:10.5607/en.2013.22.1.11

Imai, Y., Soda, M., Inoue, H., Hattori, N., Mizuno, Y., Takahashi, R., 2001. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. Cell 105, 891–902.

Imai, Y., Soda, M., Takahashi, R., 2000. Parkin Suppresses Unfolded Protein Stress-induced Cell Death through Its E3 Ubiquitin-protein Ligase Activity. J. Biol. Chem. 275, 35661–35664. doi:10.1074/jbc.C000447200

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., Berciano, J., Combarros, O., Sainz, J., 2016. Comparative blood transcriptome analysis in idiopathic and LRRK2 G2019S-associated Parkinson's disease. Neurobiol. Aging 38, 214.e1-5. doi:10.1016/j.neurobiolaging.2015.10.026

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., Berciano, J., Combarros, O., Sainz, J., 2015. Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. Neurobiol. Aging 36, 1105–1109. doi:10.1016/j.neurobiolaging.2014.10.039

Irrcher, I., Aleyasin, H., Seifert, E.L., Hewitt, S.J., Chhabra, S., Phillips, M., Lutz, A.K., Rousseaux, M.W.C., Bevilacqua, L., Jahani-Asl, A., Callaghan, S., MacLaurin, J.G., Winklhofer, K.F., Rizzu, P., Rippstein, P., Kim, R.H., Chen, C.X., Fon, E.A., Slack, R.S., Harper, M.E., McBride, H.M., Mak, T.W., Park, D.S., 2010. Loss of the Parkinson's disease-linked gene DJ-1 perturbs mitochondrial dynamics. Hum. Mol. Genet. 19, 3734–3746. doi:10.1093/hmg/ddq288

Jiang, T.-F., Zhang, Y.-J., Zhou, H.-Y., Wang, H.-M., Tian, L.-P., Liu, J., Ding, J.-Q., Chen, S.-D., 2013. Curcumin ameliorates the neurodegenerative pathology in A53T α-synuclein cell model of Parkinson's disease through the downregulation of mTOR/p70S6K signaling and the recovery of macroautophagy. J. Neuroimmune Pharmacol. Off. J. Soc. NeuroImmune Pharmacol. 8, 356–369. doi:10.1007/s11481-012-9431-7

Jornayvaz, F.R., Shulman, G.I., 2010. Regulation of mitochondrial biogenesis. Essays Biochem. 47. doi:10.1042/bse0470069

Kay, D.M., Stevens, C.F., Hamza, T.H., Montimurro, J.S., Zabetian, C.P., Factor, S.A., Samii, A., Griffith, A., Roberts, J.W., Molho, E.S., Higgins, D.S., Gancher, S., Moses, L., Zareparsi, S., Poorkaj, P., Bird, T., Nutt, J., Schellenberg, G.D., Payami, H., 2010. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. Neurology 75, 1189–1194. doi:10.1212/WNL.0b013e3181f4d832

Keeney, P.M., Xie, J., Capaldi, R.A., Bennett, J.P., 2006. Parkinson's disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled. J. Neurosci. Off. J. Soc. Neurosci. 26, 5256–5264. doi:10.1523/JNEUROSCI.0984-06.2006

Kelley, D.E., He, J., Menshikova, E.V., Ritov, V.B., 2002. Dysfunction of Mitochondria in Human Skeletal Muscle in Type 2. Diabetes 51, 2944–2950. doi:10.2337/diabetes.51.10.2944

Keyser, R.J., Lombard, D., Veikondis, R., Carr, J., Bardien, S., 2009. Analysis of exon dosage using MLPA in South African Parkinson's disease patients. neurogenetics 11, 305–312. doi:10.1007/s10048-009-0229-6

Kilarski, L.L., Pearson, J.P., Newsway, V., Majounie, E., Knipe, M.D.W., Misbahuddin, A., Chinnery, P.F., Burn, D.J., Clarke, C.E., Marion, M.-H., Lewthwaite, A.J., Nicholl, D.J., Wood, N.W., Morrison, K.E., Williams-Gray, C.H., Evans, J.R., Sawcer, S.J., Barker, R.A., Wickremaratchi, M.M., Ben-Shlomo, Y., Williams, N.M., Morris, H.R., 2012. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 27, 1522–1529. doi:10.1002/mds.25132

Kim, S.J., Sung, J.Y., Um, J.W., Hattori, N., Mizuno, Y., Tanaka, K., Paik, S.R., Kim, J., Chung, K.C., 2003. Parkin cleaves intracellular alpha-synuclein inclusions via the activation of calpain. J. Biol. Chem. 278, 41890–41899. doi:10.1074/jbc.M306017200

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., Shimizu, N., 1998a. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392, 605–608. doi:10.1038/33416

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., Shimizu, N., 1998b. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392, 605–608. doi:10.1038/33416

Klein, C., Westenberger, A., 2012. Genetics of Parkinson's Disease. Cold Spring Harb. Perspect. Med. 2. doi:10.1101/cshperspect.a008888

Ko, H.S., Kim, S.W., Sriram, S.R., Dawson, V.L., Dawson, T.M., 2006. Identification of far upstream element-binding protein-1 as an authentic Parkin substrate. J. Biol. Chem. 281, 16193–16196. doi:10.1074/jbc.C600041200

Ko, H.S., von Coelln, R., Sriram, S.R., Kim, S.W., Chung, K.K.K., Pletnikova, O., Troncoso, J., Johnson, B., Saffary, R., Goh, E.L., Song, H., Park, B.-J., Kim, M.J., Kim, S., Dawson, V.L., Dawson, T.M., 2005. Accumulation of the authentic parkin substrate aminoacyl-tRNA synthetase cofactor, p38/JTV-1, leads to catecholaminergic cell death. J. Neurosci. Off. J. Soc. Neurosci. 25, 7968–7978. doi:10.1523/JNEUROSCI.2172-05.2005

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinforma. Oxf. Engl. 23, 1289–1291. doi:10.1093/bioinformatics/btm091

Kozera, B., Rapacz, M., 2013. Reference genes in real-time PCR. J. Appl. Genet. 54, 391–406. doi:10.1007/s13353-013-0173-x

Koziorowski, D., Hoffman-Zacharska, D., Sławek, J., Szirkowiec, W., Janik, P., Bal, J., Friedman, A., 2010. Low frequency of the PARK2 gene mutations in Polish patients with the early-onset form of Parkinson disease. Parkinsonism Relat. Disord. 16, 136–138. doi:10.1016/j.parkreldis.2009.06.010

Krebiehl, G., Ruckerbauer, S., Burbulla, L.F., Kieper, N., Maurer, B., Waak, J., Wolburg, H., Gizatullina, Z., Gellerich, F.N., Woitalla, D., Riess, O., Kahle, P.J., Proikas-Cezanne, T., Krüger, R., 2010. Reduced basal autophagy and impaired mitochondrial dynamics due to loss of Parkinson's disease-associated protein DJ-1. PloS One 5, e9367. doi:10.1371/journal.pone.0009367

Krebs, C.E., Karkheiran, S., Powell, J.C., Cao, M., Makarov, V., Darvish, H., Di Paolo, G., Walker, R.H., Shahidi, G.A., Buxbaum, J.D., De Camilli, P., Yue, Z., Paisán-Ruiz, C., 2013. The Sac1 Domain of SYNJ1 Identified Mutated in a Family with Early-Onset Progressive Parkinsonism with Generalized Seizures. Hum. Mutat. 34, 1200–1207. doi:10.1002/humu.22372

Lambert, A.J., Brand, M.D., 2004. Inhibitors of the quinone-binding site allow rapid superoxide production from mitochondrial NADH:ubiquinone oxidoreductase (complex I). J. Biol. Chem. 279, 39414–39420. doi:10.1074/jbc.M406576200

Lei, P., Ayton, S., Finkelstein, D.I., Adlard, P.A., Masters, C.L., Bush, A.I., 2010. Tau protein: Relevance to Parkinson's disease. Int. J. Biochem. Cell Biol. 42, 1775–1778. doi:10.1016/j.biocel.2010.07.016

Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M.J., Jonnalagada, S., Chernova, T., Dehejia, A., Lavedan, C., Gasser, T., Steinbach, P.J., Wilkinson, K.D., Polymeropoulos, M.H., 1998. The ubiquitin pathway in Parkinson's disease. Nature 395, 451–452. doi:10.1038/26652

Lev, N., Melamed, E., Offen, D., 2003. Apoptosis and Parkinson's disease. Prog. Neuropsychopharmacol. Biol. Psychiatry, Apoptosis and Human Neurodegenerative Diseases 27, 245–250. doi:10.1016/S0278-5846(03)00019-8

Lewis, P.A., Cookson, M.R., 2012. Gene expression in the Parkinson's disease brain. Brain Res. Bull. 88, 302–312. doi:10.1016/j.brainresbull.2011.11.016

Libura, M., Pawełczyk, M., Florek, I., Matiakowska, K., Jaźwiec, B., Borg, K., Solarska, I., Zawada, M., Czekalska, S., Libura, J., Salamanczuk, Z., Jakóbczyk, M., Mucha, B., Duszeńko, E., Soszyńska, K., Karabin, K., Piątkowska-Jakubas, B., Całbecka, M., Gajkowska-Kulig, J., Gadomska, G., Kiełbiński, M., Ejduk, A., Kata, D., Grosicki, S., Kyrcz-Krzemień, S., Warzocha, K., Kuliczkowski, K., Skotnicki, A., Jęrzejczak, W.W., Haus, O., 2015. CEBPA copy number variations in normal karyotype acute myeloid leukemia: Possible role of breakpoint-associated microhomology and chromatin status in CEBPA mutagenesis. Blood Cells. Mol. Dis. 55, 284–292. doi:10.1016/j.bcmd.2015.07.002

Liew, C.-C., Ma, J., Tang, H.-C., Zheng, R., Dempsey, A.A., 2006. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J. Lab. Clin. Med. 147, 126–132. doi:10.1016/j.lab.2005.10.005

Lin, M.T., Beal, M.F., 2006. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443, 787–795. doi:10.1038/nature05292

Liu, W., Acín-Peréz, R., Geghman, K.D., Manfredi, G., Lu, B., Li, C., 2011. Pink1 regulates the oxidative phosphorylation machinery via mitochondrial fission. Proc. Natl. Acad. Sci. 108, 12920–12924. doi:10.1073/pnas.1107332108

Lockhart, D.J., Winzeler, E.A., 2000. Genomics, gene expression and DNA arrays. Nature 405, 827–836. doi:10.1038/35015701

Lohmann, E., Coquel, A.-S., Honoré, A., Gurvit, H., Hanagasi, H., Emre, M., Leutenegger, A.L., Drouet, V., Sahbatou, M., Guven, G., Erginel-Unaltuna, N., Deleuze, J.-F., Lesage, S., Brice, A., 2015. A new F-box protein 7 gene mutation causing typical Parkinson's disease. Mov. Disord. 30, 1130–1133. doi:10.1002/mds.26266

López-Lluch, G., Irusta, P.M., Navas, P., de Cabo, R., 2008. Mitochondrial biogenesis and healthy aging. Exp. Gerontol. 43, 813–819. doi:10.1016/j.exger.2008.06.014

Lowell, B.B., Shulman, G.I., 2005. Mitochondrial dysfunction and type 2 diabetes. Science 307, 384–387. doi:10.1126/science.1104343

Lu, L., Neff, F., Alvarez-Fischer, D., Henze, C., Xie, Y., Oertel, W.H., Schlegel, J., Hartmann, A., 2005. Gene expression profiling of Lewy body-bearing neurons in Parkinson's disease. Exp. Neurol. 195, 27–39. doi:10.1016/j.expneurol.2005.04.011

Lu, L., Neff, F., Fischer, D.A., Henze, C., Hirsch, E.C., Oertel, W.H., Schlegel, J., Hartmann, A., 2006. Regional vulnerability of mesencephalic dopaminergic neurons prone to degenerate in Parkinson's disease: a post-mortem study in human control subjects. Neurobiol. Dis. 23, 409–421. doi:10.1016/j.nbd.2006.04.002

Lücking, C.B., Dürr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B.S., Meco, G., Denèfle, P., Wood, N.W., Agid, Y., Nicholl, D., Breteler, M.M.B., Oostra, B.A., De Mari, M., Marconi, R., Filla, A., Bonnet, A.-M., Broussolle, E., Pollak, P., Rascol, O., Rosier, M., Arnould, A., Brice, A., 2000. Association between Early-Onset Parkinson's Disease and Mutations in the Parkin Gene. N. Engl. J. Med. 342, 1560–1567. doi:10.1056/NEJM200005253422103

Madrazo, J.A., Kelly, D.P., 2008. The PPAR trio: regulators of myocardial energy metabolism in health and disease. J. Mol. Cell. Cardiol. 44, 968–975. doi:10.1016/j.yjmcc.2008.03.021

Mandel, S., Grünblatt, E., Youdim, M., 2000. cDNA microarray to study gene expression of dopaminergic neurodegeneration and neuroprotection in MPTP and 6-hydroxydopamine models: implications for idiopathic Parkinson's disease. J. Neural Transm. Suppl. 117–124.

Margis, R., Margis, R., Rieder, C.R.M., 2011. Identification of blood microRNAs associated to Parkinsonʹs disease. J. Biotechnol. 152, 96–101. doi:10.1016/j.jbiotec.2011.01.023

Mariani, E., Polidori, M.C., Cherubini, A., Mecocci, P., 2005. Oxidative stress in brain aging, neurodegenerative and vascular diseases: An overview. J. Chromatogr. B, Analysis of Antioxidants and Biomarkers of Oxidative Stress 827, 65–75. doi:10.1016/j.jchromb.2005.04.023

Martins, M., Rosa, A., Guedes, L.C., Fonseca, B.V., Gotovac, K., Violante, S., Mestre, T., Coelho, M., Rosa, M.M., Martin, E.R., Vance, J.M., Outeiro, T.F., Wang, L., Borovecki, F., Ferreira, J.J., Oliveira, S.A., 2011. Convergence of miRNA Expression Profiling, α-Synuclein Interacton and GWAS in Parkinson's Disease. PLoS ONE 6, e25443. doi:10.1371/journal.pone.0025443

Martins-Branco, D., Esteves, A.R., Santos, D., Arduino, D.M., Swerdlow, R.H., Oliveira, C.R., Januario, C., Cardoso, S.M., 2012. Ubiquitin proteasome system in Parkinson's disease: A keeper or a witness? Exp. Neurol. 238, 89–99. doi:10.1016/j.expneurol.2012.08.008

Mata, I.F., Shi, M., Agarwal, P., Chung, K.A., Edwards, K.L., Factor, S.A., Galasko, D.R., Ginghina, C., Griffith, A., Higgins, D.S., Kay, D.M., Kim, H., Leverenz, J.B., Quinn, J.F., Roberts, J.W., Samii, A., Snapinn, K.W., Tsuang, D.W., Yearout, D., Zhang, J., Payami, H., Zabetian, C.P., 2010. SNCA variant associated with Parkinson disease and plasma alpha-synuclein level. Arch. Neurol. 67, 1350–1356. doi:10.1001/archneurol.2010.279

Matsusue, K., Gavrilova, O., Lambert, G., Brewer, H.B., Ward, J.M., Inoue, Y., LeRoith, D., Gonzalez, F.J., 2004. Hepatic CCAAT/enhancer binding protein alpha mediates induction of lipogenesis and regulation of glucose homeostasis in leptin-deficient mice. Mol. Endocrinol. Baltim. Md 18, 2751–2764. doi:10.1210/me.2004-0213

McNaught, K.S., Jenner, P., 2001. Proteasomal function is impaired in substantia nigra in Parkinson's disease. Neurosci. Lett. 297, 191–194.

McNaught, K.S.P., Belizaire, R., Isacson, O., Jenner, P., Olanow, C.W., 2003. Altered proteasomal function in sporadic Parkinson's disease. Exp. Neurol. 179, 38–46.

Mikulak, J., Bozzo, L., Roberto, A., Pontarini, E., Tentorio, P., Hudspeth, K., Lugli, E., Mavilio, D., 2014. Dopamine inhibits the effector functions of activated NK cells via the upregulation of the D5 receptor. J. Immunol. Baltim. Md 1950 193, 2792–2800. doi:10.4049/jimmunol.1401114

Miller, S.G., De Vos, P., Guerre-Millo, M., Wong, K., Hermann, T., Staels, B., Briggs, M.R., Auwerx, J., 1996. The adipocyte specific transcription factor C/EBPalpha modulates human ob gene expression. Proc. Natl. Acad. Sci. U. S. A. 93, 5507–5511.

Mizuno, Y., Ohta, S., Tanaka, M., Takamiya, S., Suzuki, K., Sato, T., Oya, H., Ozawa, T., Kagawa, Y., 1989. Deficiencies in complex I subunits of the respiratory chain in Parkinson's disease. Biochem. Biophys. Res. Commun. 163, 1450–1455.

Moisoi, N., Klupsch, K., Fedele, V., East, P., Sharma, S., Renton, A., Plun-Favreau, H., Edwards, R.E., Teismann, P., Esposti, M.D., Morrison, A.D., Wood, N.W., Downward, J., Martins, L.M., 2008. Mitochondrial dysfunction triggered by loss of HtrA2 results in the activation of a brain-specific transcriptional stress response. Cell Death Differ. 16, 449–464. doi:10.1038/cdd.2008.166

Molochnikov, L., Rabey, J.M., Dobronevsky, E., Bonucelli, U., Ceravolo, R., Frosini, D., Grünblatt, E., Riederer, P., Jacob, C., Aharon-Peretz, J., Bashenko, Y., Youdim, M.B.H., Mandel, S.A., 2012. A molecular signature in blood identifies early Parkinson's disease. Mol. Neurodegener. 7, 26. doi:10.1186/1750-1326-7-26

Moore, D.J., West, A.B., Dikeman, D.A., Dawson, V.L., Dawson, T.M., 2008. Parkin mediates the degradation-independent ubiquitination of Hsp70. J. Neurochem. 105, 1806–1819. doi:10.1111/j.1471-4159.2008.05261.x

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D., Groop, L.C., 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267–273. doi:10.1038/ng1180

Moran, L.B., Duke, D.C., Deprez, M., Dexter, D.T., Pearce, R.K.B., Graeber, M.B., 2006. Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. Neurogenetics 7, 1–11. doi:10.1007/s10048-005-0020-2

Moran, L.B., Graeber, M.B., 2008. Towards a pathway definition of Parkinson's disease: a complex disorder with links to cancer, diabetes and inflammation. Neurogenetics 9, 1–13. doi:10.1007/s10048-007-0116-y

Morrison, K.E., 2003. Parkin mutations and early onset parkinsonism. Brain 126, 1250–1251. doi:10.1093/brain/awg189

Müftüoglu, M., Elibol, B., Dalmizrak, O., Ercan, A., Kulaksiz, G., Ogüs, H., Dalkara, T., Ozer, N., 2004. Mitochondrial complex I and IV activities in leukocytes from patients with parkin mutations. Mov. Disord. Off. J. Mov. Disord. Soc. 19, 544–548. doi:10.1002/mds.10695

Mutez, E., Larvor, L., Leprêtre, F., Mouroux, V., Hamalek, D., Kerckaert, J.-P., Pérez-Tur, J., Waucquier, N., Vanbesien-Mailliot, C., Duflot, A., Devos, D., Defebvre, L., Kreisler, A., Frigard, B., Destée, A., Chartier-Harlin, M.-C., 2011. Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. Neurobiol. Aging 32, 1839–1848. doi:10.1016/j.neurobiolaging.2009.10.016

Mutez, E., Nkiliza, A., Belarbi, K., de Broucker, A., Vanbesien-Mailliot, C., Bleuse, S., Duflot, A., Comptdaer, T., Semaille, P., Blervaque, R., Hot, D., Leprêtre, F., Figeac, M., Destée, A., Chartier-Harlin, M.-C., 2014. Involvement of the immune system, endocytosis and EIF2 signaling in both genetically determined and sporadic forms of Parkinson's disease. Neurobiol. Dis. 63, 165–170. doi:10.1016/j.nbd.2013.11.007

Narendra, D., Tanaka, A., Suen, D.-F., Youle, R.J., 2009. Parkin-induced mitophagy in the pathogenesis of Parkinson disease. Autophagy 5, 706–708.

Narendra, D., Tanaka, A., Suen, D.-F., Youle, R.J., 2008. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. J. Cell Biol. 183, 795–803. doi:10.1083/jcb.200809125

Narendra, D.P., Jin, S.M., Tanaka, A., Suen, D.-F., Gautier, C.A., Shen, J., Cookson, M.R., Youle, R.J., 2010. PINK1 Is Selectively Stabilized on Impaired Mitochondria to Activate Parkin. PLoS Biol 8, e1000298. doi:10.1371/journal.pbio.1000298

Nelson, D.E., Randle, S.J., Laman, H., 2013. Beyond ubiquitination: the atypical functions of Fbxo7 and other F-box proteins. Open Biol. 3, 130131. doi:10.1098/rsob.130131

Ng, C.-H., Mok, S.Z.S., Koh, C., Ouyang, X., Fivaz, M.L., Tan, E.-K., Dawson, V.L., Dawson, T.M., Yu, F., Lim, K.-L., 2009. Parkin protects against LRRK2 G2019S mutant-induced dopaminergic neurodegeneration in Drosophila. J. Neurosci. Off. J. Soc. Neurosci. 29, 11257–11262. doi:10.1523/JNEUROSCI.2375-09.2009

Nikolova, G., Mancheva, V., 2012. Analysis of the parameters of oxidative stress in patients with Parkinson's disease. Comp. Clin. Pathol. 22, 151–155. doi:10.1007/s00580-012-1407-8

Noureddine, M.A., Li, Y.-J., van der Walt, J.M., Walters, R., Jewett, R.M., Xu, H., Wang, T., Walter, J.W., Scott, B.L., Hulette, C., Schmechel, D., Stenger, J.E., Dietrich, F., Vance, J.M., Hauser, M.A., 2005. Genomic convergence to identify candidate genes for Parkinson disease: SAGE analysis of the substantia nigra. Mov. Disord. Off. J. Mov. Disord. Soc. 20, 1299–1309. doi:10.1002/mds.20573

Paisán-Ruíz, C., Jain, S., Evans, E.W., Gilks, W.P., Simón, J., van der Brug, M., de Munain, A.L., Aparicio, S., Gil, A.M., Khan, N., Johnson, J., Martinez, J.R., Nicholl, D., Carrera, I.M., Peña, A.S., de Silva, R., Lees, A., Martí-Massó, J.F., Pérez-Tur, J., Wood, N.W., Singleton, A.B., 2004. Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease. Neuron 44, 595–600. doi:10.1016/j.neuron.2004.10.023

Palacino, J.J., Sagi, D., Goldberg, M.S., Krauss, S., Motz, C., Wacker, M., Klose, J., Shen, J., 2004. Mitochondrial Dysfunction and Oxidative Damage in parkin-deficient Mice. J. Biol. Chem. 279, 18614–18622. doi:10.1074/jbc.M401135200

Pankratz, N., Dumitriu, A., Hetrick, K.N., Sun, M., Latourelle, J.C., Wilk, J.B., Halter, C., Doheny, K.F., Gusella, J.F., Nichols, W.C., Myers, R.H., Foroud, T., DeStefano, A.L., Investigators, the P.-P. and

G., Laboratories, C. and M.G., 2011. Copy Number Variation in Familial Parkinson Disease. PLOS ONE 6, e20988. doi:10.1371/journal.pone.0020988

Pankratz, N., Kissell, D.K., Pauciulo, M.W., Halter, C.A., Rudolph, A., Pfeiffer, R.F., Marder, K.S., Foroud, T., Nichols, W.C., 2009. Parkin dosage mutations have greater pathogenicity in familial PD than simple sequence mutations. Neurology 73, 279–286. doi:10.1212/WNL.0b013e3181af7a33

Parker, W.D., Parks, J.K., Swerdlow, R.H., 2008. Complex I Deficiency in Parkinson's Disease Frontal Cortex. Brain Res. 1189, 215–218. doi:10.1016/j.brainres.2007.10.061

Patti, M.E., Butte, A.J., Crunkhorn, S., Cusi, K., Berria, R., Kashyap, S., Miyazaki, Y., Kohane, I., Costello, M., Saccone, R., Landaker, E.J., Goldfine, A.B., Mun, E., DeFronzo, R., Finlayson, J., Kahn, C.R., Mandarino, L.J., 2003. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. Proc. Natl. Acad. Sci. 100, 8466–8471. doi:10.1073/pnas.1032913100

Periquet, M., Lücking, C., Vaughan, J., Bonifati, V., Dürr, A., De Michele, G., Horstink, M., Farrer, M., Illarioshkin, S.N., Pollak, P., Borg, M., Brefel-Courbon, C., Denefle, P., Meco, G., Gasser, T., Breteler, M.M., Wood, N., Agid, Y., Brice, A., French Parkinson's Disease Genetics Study Group. The European Consortium on Genetic Susceptibility in Parkinson's Disease, 2001. Origin of the mutations in the parkin gene in Europe: exon rearrangements are independent recurrent events, whereas point mutations may result from Founder effects. Am. J. Hum. Genet. 68, 617–626.

Petersen, K.F., Dufour, S., Befroy, D., Garcia, R., Shulman, G.I., 2004. Impaired Mitochondrial Activity in the Insulin-Resistant Offspring of Patients with Type 2 Diabetes. N. Engl. J. Med. 350, 664–671. doi:10.1056/NEJMoa031314

Petit, A., Kawarai, T., Paitel, E., Sanjo, N., Maj, M., Scheid, M., Chen, F., Gu, Y., Hasegawa, H., Salehi-Rad, S., Wang, L., Rogaeva, E., Fraser, P., Robinson, B., St George-Hyslop, P., Tandon, A., 2005. Wild-type PINK1 prevents basal and induced neuronal apoptosis, a protective effect abrogated by Parkinson disease-related mutations. J. Biol. Chem. 280, 34025–34032. doi:10.1074/jbc.M505143200

Pfaffl, M.W., Horgan, G.W., Dempfle, L., 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 30, e36.

Plun-Favreau, H., Hardy, J., 2008. PINK1 in mitochondrial function. Proc. Natl. Acad. Sci. 105, 11041–11042. doi:10.1073/pnas.0805908105

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E.S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W.G., Lazzarini, A.M., Duvoisin, R.C., Di Iorio, G., Golbe, L.I., Nussbaum, R.L., 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276, 2045–2047.

Puigserver, P., Spiegelman, B.M., 2003. Peroxisome Proliferator-Activated Receptor-γ Coactivator 1α (PGC-1α): Transcriptional Coactivator and Metabolic Regulator. Endocr. Rev. 24, 78–90. doi:10.1210/er.2002-0012

Quadri, M., Fang, M., Picillo, M., Olgiati, S., Breedveld, G.J., Graafland, J., Wu, B., Xu, F., Erro, R., Amboni, M., Pappatà, S., Quarantelli, M., Annesi, G., Quattrone, A., Chien, H.F., Barbosa, E.R., The International Parkinsonism Genetics Network, Oostra, B.A., Barone, P., Wang, J., Bonifati, V., 2013. Mutation in the SYNJ1 Gene Associated with Autosomal Recessive, Early-Onset Parkinsonism. Hum. Mutat. 34, 1208–1215. doi:10.1002/humu.22373

Ramirez, A., Heimbach, A., Gründemann, J., Stiller, B., Hampshire, D., Cid, L.P., Goebel, I., Mubaidin, A.F., Wriekat, A.-L., Roeper, J., Al-Din, A., Hillmer, A.M., Karsak, M., Liss, B., Woods, C.G., Behrens, M.I., Kubisch, C., 2006. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nat. Genet. 38, 1184–1191. doi:10.1038/ng1884

Reddy, S.D.N., Pakala, S.B., Ohshiro, K., Rayala, S.K., Kumar, R., 2009. MicroRNA-661, a c/EBPalpha target, inhibits metastatic tumor antigen 1 and regulates its functions. Cancer Res. 69, 5639–5642. doi:10.1158/0008-5472.CAN-09-0898

Riley, B.E., Gardai, S.J., Emig-Agius, D., Bessarabova, M., Ivliev, A.E., Schüle, B., Schüle, B., Alexander, J., Wallace, W., Halliday, G.M., Langston, J.W., Braxton, S., Yednock, T., Shaler, T., Johnston, J.A., 2014. Systems-based analyses of brain regions functionally impacted in Parkinson's disease reveals underlying causal mechanisms. PloS One 9, e102909. doi:10.1371/journal.pone.0102909

Sandyk, R., 1993. The relationship between diabetes mellitus and Parkinson's disease. Int. J. Neurosci. 69, 125–130.

Santiago, J.A., Potashkin, J.A., 2015. Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. 112, 2257–2262. doi:10.1073/pnas.1423573112

Santiago, J.A., Scherzer, C.R., Harvard Biomarker Study, Potashkin, J.A., 2013. Specific splice variants are associated with Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. 28, 1724–1727. doi:10.1002/mds.25635

Sargent, C.A., Boucher, C.A., Kirsch, S., Brown, G., Weiss, B., Trundley, A., Burgoyne, P., Saut, N., Durand, C., Levy, N., Terriou, P., Hargreave, T., Cooke, H., Mitchell, M., Rappold, G.A., Affara, N.A., 1999. The critical region of overlap defining the AZFa male infertility interval of proximal Yq contains three transcribed sequences. J. Med. Genet. 36, 670–677.

Sarojam, S., Raveendran, S., Vijay, S., Sreedharan, J., Narayanan, G., Sreedharan, H., 2015. Characterization of CEBPA Mutations and Polymorphisms and their Prognostic Relevance in De Novo Acute Myeloid Leukemia Patients. Asian Pac. J. Cancer Prev. APJCP 16, 3785–3792.

Schapira, A.H.V., Cooper, J.M., Dexter, D., Jenner, P., Clark, J.B., Marsden, C.D., 1989. MITOCHONDRIAL COMPLEX I DEFICIENCY IN PARKINSON'S DISEASE. The Lancet, Originally published as Volume 1, Issue 8649 333, 1269. doi:10.1016/S0140-6736(89)92366-0

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., Sudarsky, L.R., Standaert, D.G., Growdon, J.H., Jensen, R.V., Gullans, S.R., 2007a. Molecular markers of early Parkinson's disease based on gene expression in blood. Proc. Natl. Acad. Sci. U. S. A. 104, 955–960. doi:10.1073/pnas.0610204104

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., Sudarsky, L.R., Standaert, D.G., Growdon, J.H., Jensen, R.V., Gullans, S.R., 2007b. Molecular markers of early Parkinson's disease based on gene expression in blood. Proc. Natl. Acad. Sci. U. S. A. 104, 955–960. doi:10.1073/pnas.0610204104

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., Ragg, T., 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. 7, 3. doi:10.1186/1471-2199-7-3

Schulze, A., Downward, J., 2001. Navigating gene expression using microarrays--a technology review. Nat. Cell Biol. 3, E190-195. doi:10.1038/35087138

Sellbach, A.N., Boyle, R.S., Silburn, P.A., Mellick, G.D., 2006. Parkinson's disease and family history. Parkinsonism Relat. Disord. 12, 399–409. doi:10.1016/j.parkreldis.2006.03.002

Serafin, A., Foco, L., Blankenburg, H., Picard, A., Zanigni, S., Zanon, A., Pramstaller, P.P., Hicks, A.A., Schwienbacher, C., 2014. Identification of a set of endogenous reference genes for miRNA expression studies in Parkinson's disease blood samples. BMC Res. Notes 7, 715. doi:10.1186/1756-0500-7-715

Sherer, T.B., Betarbet, R., Testa, C.M., Seo, B.B., Richardson, J.R., Kim, J.H., Miller, G.W., Yagi, T., Matsuno-Yagi, A., Greenamyre, J.T., 2003a. Mechanism of toxicity in rotenone models of Parkinson's disease. J. Neurosci. Off. J. Soc. Neurosci. 23, 10756–10764.

Sherer, T.B., Kim, J.-H., Betarbet, R., Greenamyre, J.T., 2003b. Subcutaneous Rotenone Exposure Causes Highly Selective Dopaminergic Degeneration and α-Synuclein Aggregation. Exp. Neurol. 179, 9–16. doi:10.1006/exnr.2002.8072

Shi, C.-H., Mao, C.-Y., Zhang, S.-Y., Yang, J., Song, B., Wu, P., Zuo, C.-T., Liu, Y.-T., Ji, Y., Yang, Z.-H., Wu, J., Zhuang, Z.-P., Xu, Y.-M., 2016. CHCHD2 gene mutations in familial and sporadic Parkinson's disease. Neurobiol. Aging 38, 217.e9-13. doi:10.1016/j.neurobiolaging.2015.10.040

Shimura, H., Hattori, N., Kubo, S. i, Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., Suzuki, T., 2000. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat. Genet. 25, 302–305. doi:10.1038/77060

Shojaee, S., Sina, F., Banihosseini, S.S., Kazemi, M.H., Kalhor, R., Shahidi, G.-A., Fakhrai-Rad, H., Ronaghi, M., Elahi, E., 2008. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. Am. J. Hum. Genet. 82, 1375–1384. doi:10.1016/j.ajhg.2008.05.005

Sian, J., Youdim, M.B.H., Riederer, P., Gerlach, M., 1999a. MPTP-Induced Parkinsonian Syndrome [WWW Document]. URL http://www.ncbi.nlm.nih.gov/books/NBK27974/ (accessed 6.15.14).

Simunovic, F., Yi, M., Wang, Y., Macey, L., Brown, L.T., Krichevsky, A.M., Andersen, S.L., Stephens, R.M., Benes, F.M., Sonntag, K.C., 2009. Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. Brain 132, 1795–1809. doi:10.1093/brain/awn323

Simunovic, F., Yi, M., Wang, Y., Stephens, R., Sonntag, K.C., 2010. Evidence for Gender-Specific Transcriptional Profiles of Nigral Dopamine Neurons in Parkinson Disease. PLoS ONE 5, e8856. doi:10.1371/journal.pone.0008856

Singh, A., Kumar, N., 2013. A REVIEW ON DNA MICROARRAY TECHNOLOGY. Int. J. Curr. Res. Rev. 5, 01–05.

Smith, M.L., Cavenagh, J.D., Lister, T.A., Fitzgibbon, J., 2004. Mutation of CEBPA in familial acute myeloid leukemia. N. Engl. J. Med. 351, 2403–2407. doi:10.1056/NEJMoa041331

Soneson, C., Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14, 91. doi:10.1186/1471-2105-14-91

Soreq, L., Bergman, H., Israel, Z., Soreq, H., 2012. Exon arrays reveal alternative splicing aberrations in Parkinson's disease leukocytes. Neurodegener. Dis. 10, 203–206. doi:10.1159/000332598

Soreq, L., Guffanti, A., Salomonis, N., Simchovitz, A., Israel, Z., Bergman, H., Soreq, H., 2014. Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. PLoS Comput. Biol. 10, e1003517. doi:10.1371/journal.pcbi.1003517

Soreq, L., Israel, Z., Bergman, H., Soreq, H., 2008. Advanced microarray analysis highlights modified neuro-immune signaling in nucleated blood cells from Parkinson's disease patients. J. Neuroimmunol. 201–202, 227–236. doi:10.1016/j.jneuroim.2008.06.019

Spillantini, M.G., Schmidt, M.L., Lee, V.M.-Y., Trojanowski, J.Q., Jakes, R., Goedert, M., 1997. α-Synuclein in Lewy bodies. Nature 388, 839–840. doi:10.1038/42166

Spiridon Papapetropoulos, J.F.-M., 2006. Multiregional Gene Expression Profiling Identifies MRPS6 as a Possible Candidate Gene for Parkinson's Disease. Gene Expr. 13, 205–15. doi:10.3727/000000006783991827

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J., 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82. doi:10.1126/science.aaf2403

Stamper, C., Siegel, A., Liang, W.S., Pearson, J.V., Stephan, D.A., Shill, H., Connor, D., Caviness, J.N., Sabbagh, M., Beach, T.G., Adler, C.H., Dunckley, T., 2008. Neuronal gene expression correlates of Parkinson's disease with dementia. Mov. Disord. Off. J. Mov. Disord. Soc. 23, 1588–1595. doi:10.1002/mds.22184

Staropoli, J.F., McDermott, C., Martinat, C., Schulman, B., Demireva, E., Abeliovich, A., 2003. Parkin is a component of an SCF-like ubiquitin ligase complex and protects postmitotic neurons from kainate excitotoxicity. Neuron 37, 735–749.

Subramaniam, M., Althof, D., Gispert, S., Schwenk, J., Auburger, G., Kulik, A., Fakler, B., Roeper, J., 2014. Mutant α-Synuclein Enhances Firing Frequencies in Dopamine Substantia Nigra Neurons by Oxidative Impairment of A-Type Potassium Channels. J. Neurosci. Off. J. Soc. Neurosci. 34, 13586–13599. doi:10.1523/JNEUROSCI.5069-13.2014

Sullivan, P.F., Fan, C., Perou, C.M., 2006. Evaluating the comparability of gene expression in blood and brain. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 141B, 261–268. doi:10.1002/ajmg.b.30272

Taira, T., Saito, Y., Niki, T., Iguchi-Ariga, S.M.M., Takahashi, K., Ariga, H., 2004. DJ-1 has a role in antioxidative stress to prevent cell death. EMBO Rep. 5, 213–218. doi:10.1038/sj.embor.7400074

Tan, E.K., Shen, H., Tan, J.M.M., Lim, K.L., Fook-Chong, S., Hu, W.P., Paterson, M.C., Chandran, V.R., Yew, K., Tan, C., Yuen, Y., Pavanni, R., Wong, M.C., Puvan, K., Zhao, Y., 2005. Differential expression of splice variant and wild-type parkin in sporadic Parkinson's disease. Neurogenetics 6, 179–184. doi:10.1007/s10048-005-0001-5

Tsunemi, T., Krainc, D., 2014. $Zn^{2+}$ dyshomeostasis caused by loss of ATP13A2/PARK9 leads to lysosomal dysfunction and alpha-synuclein accumulation. Hum. Mol. Genet. 23, 2791–2801. doi:10.1093/hmg/ddt572

Uversky, V.N., 2004. Neurotoxicant-induced animal models of Parkinson?s disease: understanding the role of rotenone, maneb and paraquat in neurodegeneration. Cell Tissue Res. 318, 225–241. doi:10.1007/s00441-004-0937-z

Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M.K., Harvey, K., Gispert, S., Ali, Z., Turco, D.D., Bentivoglio, A.R., Healy, D.G., Albanese, A., Nussbaum, R., González-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W.P., Latchman, D.S., Harvey, R.J., Dallapiccola, B., Auburger, G., Wood, N.W., 2004. Hereditary Early-Onset Parkinson's Disease Caused by Mutations in PINK1. Science 304, 1158–1160. doi:10.1126/science.1096284

van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., Radhakrishnan, A., Tan, S.-T., Voss, K., Weichenberger, C.X., Albers, C.A., Al-Hussani, A., Asselbergs, F.W., Ciullo, M., Danjou, F., Dina, C., Esko, T., Evans, D.M., Franke, L., Gögele, M., Hartiala, J., Hersch, M., Holm, H., Hottenga, J.-J., Kanoni, S., Kleber, M.E., Lagou, V., Langenberg, C., Lopez, L.M., Lyytikäinen, L.-P., Melander, O., Murgia, F., Nolte, I.M., O'Reilly, P.F., Padmanabhan, S., Parsa, A., Pirastu, N., Porcu, E., Portas, L., Prokopenko, I., Ried, J.S., Shin, S.-Y., Tang, C.S., Teumer, A., Traglia, M., Ulivi, S., Westra, H.-J., Yang, J., Hua Zhao, J., Anni, F., Abdellaoui, A., Attwood, A., Balkau, B., Bandinelli, S., Bastardot, F., Benyamin, B., Boehm, B.O., Cookson, W.O., Das, D., de Bakker, P.I.W., de Boer, R.A., de Geus, E.J.C., de Moor, M.H., Dimitriou, M., Domingues, F.S., Döring, A., Engström, G., Ingi Eyjolfsson, G., Ferrucci, L., Fischer, K., Galanello, R., Garner, S.F., Genser, B., Gibson, Q.D., Girotto, G., Fannar Gudbjartsson, D., Harris, S.E., Hartikainen, A.-L., Hastie, C.E., Hedblad, B., Illig, T., Jolley, J., Kähönen, M., Kema, I.P., Kemp, J.P., Liang, L., Lloyd-Jones, H., Loos, R.J.F., Meacham, S., Medland, S.E., Meisinger, C., Memari, Y., Mihailov, E., Miller, K., Moffatt, M.F., Nauck, M., Novatchkova, M., Nutile, T., Olafsson, I., Onundarson, P.T., Parracciani, D., Penninx, B.W., Perseu, L., Piga, A., Pistis, G., Pouta, A., Puc, U., Raitakari, O., Ring, S.M., Robino, A., Ruggiero, D., Ruokonen, A., Saint-Pierre, A., Sala, C., Salumets, A., Sambrook, J., Schepers, H., Oliver Schmidt, C., Silljé, H.H.W., Sladek, R., Smit, J.H., Starr, J.M., Stephens, J., Sulem, P., Tanaka, T., Thorsteinsdottir, U., Tragante, V., van Gilst, W.H., Joost van Pelt, L., van Veldhuisen, D.J., Völker, U., Whitfield, J.B., Willemsen, G., Winkelmann, B.R., Wirnsberger, G., Algra, A., Cucca, F., d'Adamo, A.P., Danesh, J., Deary, I.J., Dominiczak, A.F., Elliott, P., Fortina, P., Froguel, P., Gasparini, P., Greinacher, A., Hazen, S.L., Jarvelin, M.-R., Tee Khaw, K., Lehtimäki, T., Maerz, W., Martin, N.G., Metspalu, A., Mitchell, B.D., Montgomery, G.W., Moore, C., Navis, G., Pirastu, M., Pramstaller, P.P., Ramirez-Solis, R., Schadt, E., Scott, J., Shuldiner, A.R., Davey Smith, G., Gustav Smith, J., Snieder, H., Sorice, R., Spector, T.D., Stefansson, K., Stumvoll, M., Wilson Tang, W.H., Toniolo, D., Tönjes, A., Visscher, P.M., Vollenweider, P., Wareham, N.J., Wolffenbuttel, B.H.R., Boomsma, D.I., Beckmann, J.S.,

Dedoussis, G.V., Deloukas, P., Ferreira, M.A., Sanna, S., Uda, M., Hicks, A.A., Martin Penninger, J., Gieger, C., Kooner, J.S., Ouwehand, W.H., Soranzo, N., Chambers, J.C., 2012. Seventy-five genetic loci influencing the human red blood cell. Nature 492, 369–375. doi:10.1038/nature11677

van der Merwe, C., Carr, J., Glanzmann, B., Bardien, S., 2016. Exonic rearrangements in the known Parkinson's disease-causing genes are a rare cause of the disease in South African patients. Neurosci. Lett. 619, 168–171. doi:10.1016/j.neulet.2016.03.028

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.

Vilarino-Guell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., Behrouz, B., Melrose, H.L., Hentati, E., Puschmann, A., Evans, D.M., Conibear, E., Wasserman, W.W., Aasly, J.O., Burkhard, P.R., Djaldetti, R., Ghika, J., Hentati, F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A.H., Solida, A., Wu, R.-M., Uitti, R.J., Wszolek, Z.K., Vingerhoets, F., Farrer, M.J., 2011. VPS35 Mutations in Parkinson Disease. Am. J. Hum. Genet. 89, 162–167. doi:10.1016/j.ajhg.2011.06.001

Vogt, I.R., Lees, A.J., Evert, B.O., Klockgether, T., Bonin, M., Wüllner, U., 2006. Transcriptional changes in multiple system atrophy and Parkinson's disease putamen. Exp. Neurol. 199, 465–478. doi:10.1016/j.expneurol.2006.01.008

Volpicelli-Daley, L.A., Gamble, K.L., Schultheiss, C.E., Riddle, D.M., West, A.B., Lee, V.M.-Y., 2014. Formation of α-synuclein Lewy neurite–like aggregates in axons impedes the transport of distinct endosomes. Mol. Biol. Cell 25, 4010–4023. doi:10.1091/mbc.E14-02-0741

Wang, C., Lu, R., Ouyang, X., Ho, M.W.L., Chia, W., Yu, F., Lim, K.-L., 2007. Drosophila Overexpressing Parkin R275W Mutant Exhibits Dopaminergic Neuron Degeneration and Mitochondrial Abnormalities. J. Neurosci. 27, 8563–8570. doi:10.1523/JNEUROSCI.0218-07.2007

Wang, H., Iakova, P., Wilde, M., Welm, A., Goode, T., Roesler, W.J., Timchenko, N.A., 2001. C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. Mol. Cell 8, 817–828.

Wang, Q., Liu, Y., Zhou, J., 2015. Neuroinflammation in Parkinson's disease and its potential as therapeutic target. Transl. Neurodegener. 4. doi:10.1186/s40035-015-0042-0

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. doi:10.1038/nrg2484

Wenz, T., 2009. PGC-1alpha activation as a therapeutic approach in mitochondrial disease. IUBMB Life 61, 1051–1062. doi:10.1002/iub.261

West, A.B., Moore, D.J., Biskup, S., Bugayenko, A., Smith, W.W., Ross, C.A., Dawson, V.L., Dawson, T.M., 2005. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. Proc. Natl. Acad. Sci. U. S. A. 102, 16842–16847. doi:10.1073/pnas.0507360102

WHO | Neurological Disorders: Public Health Challenges [WWW Document], n.d. . WHO. URL http://www.who.int/mental_health/neurology/neurodiso/en/ (accessed 5.26.16).

Wooten, G.F., Currie, L.J., Bovbjerg, V.E., Lee, J.K., Patrie, J., 2004. Are men at greater risk for Parkinson's disease than women? J. Neurol. Neurosurg. Psychiatry 75, 637–639. doi:10.1136/jnnp.2003.020982

Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., Quake, S.R., 2014. Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods 11, 41–46. doi:10.1038/nmeth.2694

Xia, R., Mao, Z.-H., 2012. Progression of motor symptoms in Parkinson's disease. Neurosci. Bull. 28, 39–48. doi:10.1007/s12264-012-1050-z

Xing, Y., Yu, T., Wu, Y.N., Roy, M., Kim, J., Lee, C., 2006. An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. Nucleic Acids Res. 34, 3150–3160. doi:10.1093/nar/gkl396

Yanagida, T., Tsushima, J., Kitamura, Y., Yanagisawa, D., Takata, K., Shibaike, T., Yamamoto, A., Taniguchi, T., Yasui, H., Taira, T., Morikawa, S., Inubushi, T., Tooyama, I., Ariga, H., 2009.

Oxidative stress induction of DJ-1 protein in reactive astrocytes scavenges free radicals and reduces cell injury. Oxid. Med. Cell. Longev. 2, 36–42.

Yokota, T., Sugawara, K., Ito, K., Takahashi, R., Ariga, H., Mizusawa, H., 2003. Down regulation of DJ-1 enhances cell death by oxidative stress, ER stress, and proteasome inhibition. Biochem. Biophys. Res. Commun. 312, 1342–1348. doi:10.1016/j.bbrc.2003.11.056

Zhang, J., Fitsanakis, V.A., Gu, G., Jing, D., Ao, M., Amarnath, V., Montine, T.J., 2003. Manganese ethylene-bis-dithiocarbamate and selective dopaminergic neurodegeneration in rat: a link through mitochondrial dysfunction. J. Neurochem. 84, 336–346.

Zhang, Y., Gao, J., Chung, K.K., Huang, H., Dawson, V.L., Dawson, T.M., 2000. Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. Proc. Natl. Acad. Sci. U. S. A. 97, 13354–13359. doi:10.1073/pnas.240347797

Zhang, Y., James, M., Middleton, F.A., Davis, R.L., 2005. Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 137B, 5–16. doi:10.1002/ajmg.b.30195

Zheng, B., Liao, Z., Locascio, J.J., Lesniak, K.A., Roderick, S.S., Watt, M.L., Eklund, A.C., Zhang-James, Y., Kim, P.D., Hauser, M.A., Grünblatt, E., Moran, L.B., Mandel, S.A., Riederer, P., Miller, R.M., Federoff, H.J., Wüllner, U., Papapetropoulos, S., Youdim, M.B., Cantuti-Castelvetri, I., Young, A.B., Vance, J.M., Davis, R.L., Hedreen, J.C., Adler, C.H., Beach, T.G., Graeber, M.B., Middleton, F.A., Rochet, J.-C., Scherzer, C.R., Global PD Gene Expression (GPEX) Consortium, 2010. PGC-1α, a potential therapeutic target for early intervention in Parkinson's disease. Sci. Transl. Med. 2, 52ra73. doi:10.1126/scitranslmed.3001059

Zhou, Z.D., Sathiyamoorthy, S., Angeles, D.C., Tan, E.K., 2016. Linking F-box protein 7 and parkin to neuronal degeneration in Parkinson's disease (PD). Mol. Brain 9, 41. doi:10.1186/s13041-016-0218-2

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., Rossle, S.C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Brücke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T., Strom, T.M., 2011. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am. J. Hum. Genet. 89, 168–175. doi:10.1016/j.ajhg.2011.06.008

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., Stoessl, A.J., Pfeiffer, R.F., Patenge, N., Carbajal, I.C., Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D.W., Meitinger, T., Strom, T.M., Wszolek, Z.K., Gasser, T., 2004. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44, 601–607. doi:10.1016/j.neuron.2004.11.005

## Online References:

- DAVID Bioinformatics Database: https://david.ncifcrf.gov/
- Ensembl Genome Browser database: http://www.ensembl.org
- GeneCards®: http://www.genecards.org/
- Gemma database: http://www.chibi.ubc.ca/Gemma/home.html
- Gene ontology tool: www.geneontology.org
- IDT OligoAnalyzer® software : http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer

- Ingenuity pathway analysis: www.ingenuity.com
- NCBI BLAST: http://www.ncbi.nlm.nih.gov/BLAST
- NCBI Genbank database: http://www.ncbi.nlm.nih.gov/Entrez
- National center for biotechnology information (NCBI): http://www.ncbi.nlm.nih.gov/geo/
- Online Mendelian Inheritance in Man (OMIM): http://www.omim.org/
- Partek® Flow®: http://www.partek.com/partekflow
- Ref genes database: http://www.refgenes.org/rg/

# Appendix I

## List of differentially expressed genes in blood

*Avaliable from author

# Appendix II

## List of differentially expressed genes in brain tissue

*Avaliable from author

# Appendix III

## Excluded Studies

**Table 3: Studies excluded from the genome-wide gene expression table and analysis**

| First author | Year of Publication | Reason for exclusion |
|---|---|---|
| Hauser | 2003 | No PD samples; only 2 control brain samples; SAGE |
| Lu | 2005 | Not genome-wide; RT-PCR study |
| Nourredine | 2005 | Only 3 PD and 2 control samples; SAGE as follow-up for GWAS |
| Tan | 2005 | Not genome-wide; RT-PCR study |
| Lu | 2006 | No PD samples |
| Moran | 2006 | Data were re-analyzed in Moran et al. 2008 |
| Cantuti-Castelvetri | 2007 | LCM samples; Comparison between males and females, not with PD vs. control |
| Langerveld | 2007 | No PD samples |
| Fuchs | 2008 | Not genome-wide; RT-PCR study |
| Le | 2008 | Not genome-wide; RT-PCR study |
| Simunovic | 2009 | Data were re-analyzed in Simunovic et al. 2010 |
| Margis | 2011 | Not genome-wide; RT-PCR study |
| Molochnikov | 2012 | Not genome-wide; RT-PCR study |
| Saunders | 2012 | Not genome-wide; RT-PCR study |
| Ai | 2013 | Not genome-wide; RT-PCR study |
| Santiago | 2013 | Not genome-wide; RT-PCR study |
| Trabzuni | 2013 | No PD samples |
| Alieva | 2014 | Only 4 PD and 4 control samples; pooled RNA |
| Nalls | 2014 | SNP and eQTL study; no comparison between PD and control |
| Corradini | 2014 | No gene list provided |
| Serafin | 2014 | Not genome-wide; RT-PCR study |
| Infante | 2015 | Same samples as in Infante et al. 2016 |
| Montarolo | 2016 | Not genome-wide; RT-PCR study |
| Yalcinkaya | 2016 | Not genome-wide; RT-PCR study |
| Yilmaz | 2016 | Not genome-wide; RT-PCR study |

# Appendix IV

## miRNA Expression Studies

Table 4: **Genome-wide miRNA expression studies performed on blood and brain tissue from patients with Parkinson's disease**

| miRNA/lncRNA | Chatterjee 2014: PBMC; GSE16658 | Soreq 2014: blood leukocytes | Soreq 2014: brain tissue | Martins 2011: PBMC | Botta-Orfila 2014: serum; IPDvs.CTL | Botta-Orfila 2014: serum; LRRK2-PDvs.CTL | Grand Total |
|---|---|---|---|---|---|---|---|
| AC000099.1 | | | Up | | | | |
| AC002076.10 | | | Up | | | | |
| AC003092.1 | | | Up | | | | |
| AC004744.3 | | Down | | | | | |
| AC005027.3 | | | Up | | | | |
| AC005042.4 | | | Up | | | | |
| AC005083.1 | | | Up | | | | |
| AC005522.7 | | | Down | | | | |
| AC007078.4 | | | Down | | | | |
| AC007349.4 | | | Down | | | | |
| AC007652.1 | | | Up | | | | |
| AC007879.3 | | | Up | | | | |
| AC009158.1 | | | Down | | | | |
| AC009227.2 | | | Up | | | | |
| AC012613.2 | | | Up | | | | |
| AC016723.4 | | | Up | | | | |
| AC018470.4 | | | Up | | | | |
| AC019100.3 | | | Up | | | | |
| AC019117.2 | | | Down | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AC068610.3 | | | Up | | | | |
| AC078883.4 | | | Down | | | | |
| AC079586.1 | | | Up | | | | |
| AC091133.1 | | | Up | | | | |
| AC092573.2 | | | Up | | | | |
| AC093326.3 | | | Up | | | | |
| AC103881.1 | | | Up | | | | |
| AC105393.1 | | | Up | | | | |
| AC107070.1 | | | Down | | | | |
| AC131097.3 | | | Up | | | | |
| AF196970.3 | | | Up | | | | |
| AF212831.2 | | | Up | | | | |
| AP000459.7 | | | Up | | | | |
| AP000472.3 | | | Up | | | | |
| AP000936.4 | | | Up | | | | |
| ARHGEF7-IT1 | | | Up | | | | |
| ARL5B-AS1 | | | Down | | | | |
| C17orf103 | | | Up | | | | |
| C6orf183 | | | Up | | | | |
| CCDC144B | | | Up | | | | |
| CTA-392C11.2 | | | Up | | | | |
| CTB-118N6.2 | | | Up | | | | |
| CTB-73N10.1 | | | Down | | | | |
| CTC-552D5.1 | | | Up | | | | |
| CTD-2037K23.2 | | | Down | | | | |
| CTD-2134P3.1 | | | Up | | | | |
| CTD-2151A2.1 | | | Up | | | | |
| CTD-2215L10.1 | | | Up | | | | |
| CTD-2256P15.1 | | | Down | | | | |
| CTD-2269F5.1 | | | Up | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CTD-2501M5.1 | | | Up | | | | |
| CTD-2544H17.1 | | | Up | | | | |
| CTD-2562J17.2 | | | Down | | | | |
| CWC15 | | | Down | | | | |
| ENO1-IT1 | | | Up | | | | |
| FER1L6-AS2 | | | Up | | | | |
| GS1-519E5.1 | | | Up | | | | |
| HCG21 | | | Up | | | | |
| HCG25 | | | Up | | | | |
| HLTF-AS1 | | | Up | | | | |
| HNRNPA1P57 | | | Up | | | | |
| HOXB-AS1 | | | Up | | | | |
| HOXB-AS3 | | | Down | | | | |
| hsa_SNORD118 | 1 | | | | | | |
| hsa-let-7a | 1 | | | | | | |
| hsa-let-7b | 1 | | | | | | |
| hsa-let-7c | 1 | | | | | | |
| hsa-let-7d | 1 | | | | | | |
| hsa-let-7e | 1 | | | | | | |
| hsa-let-7f | 1 | | | | | | |
| hsa-let-7f-1 | 1 | | | | | | |
| hsa-let-7g | 1 | | | | | | |
| hsa-let-7i | 1 | | | | | | |
| hsa-miR-100 | 1 | | | | Up | | 2 |
| hsa-miR-101 | 1 | | | | | | |
| hsa-miR-103 | 1 | | | | | | |
| hsa-miR-105 | 1 | | | | | | |
| hsa-miR-106a | 1 | | | | | | |
| hsa-miR-106b | 1 | | | | | | |
| hsa-miR-107 | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-10a | 1 | | | | | | |
| hsa-miR-125a-5p | 1 | | | | | | |
| hsa-miR-125b | 1 | | | | | | |
| hsa-miR-126 | 1 | | | Down | | | 2 |
| hsa-miR-126* | | | | Down | | | |
| hsa-miR-128 | 1 | | | | | | |
| hsa-miR-130a | 1 | | | | | | |
| hsa-miR-130b | 1 | | | | | | |
| hsa-miR-134 | | | | | Up | | |
| hsa-miR-138 | 1 | | | | | | |
| hsa-miR-138-1 | 1 | | | | | | |
| hsa-miR-140-3p | 1 | | | | | | |
| hsa-miR-140-5p | 1 | | | | | | |
| hsa-miR-141 | 1 | | | | | | |
| hsa-miR-142-3p | 1 | | | | | | |
| hsa-miR-142-5p | 1 | | | | | | |
| hsa-miR-143 | 1 | | | | | | |
| hsa-miR-144 | 1 | | | | | | |
| hsa-miR-145 | 1 | | | | | | |
| hsa-miR-146a | 1 | | | | | | |
| hsa-miR-146b-5p | 1 | | | | | | |
| hsa-miR-147 | 1 | | | Down | | | 2 |
| hsa-miR-148a | 1 | | | | | | |
| hsa-miR-148b | 1 | | | | | | |
| hsa-miR-149 | 1 | | | | | | |
| hsa-miR-150 | 1 | | | | | | |
| hsa-miR-151-3p | 1 | | | Down | | | 2 |
| hsa-miR-151-5p | 1 | | | Down | | | 2 |
| hsa-miR-152 | 1 | | | | | | |
| hsa-miR-153 | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-154 | 1 | | | | | | |
| hsa-miR-155 | 1 | | | | | | |
| hsa-miR-15a | 1 | | | | | | |
| hsa-miR-16 | 1 | | | | Down | Down | 3 |
| hsa-miR-16-1 | 1 | | | | | | |
| hsa-miR-16-2 | 1 | | | | | | |
| hsa-miR-17 | 1 | | | | | | |
| hsa-miR-181a | 1 | | | | Up | | 2 |
| hsa-miR-181a-2 | 1 | | | | | | |
| hsa-miR-181b | 1 | | | | | | |
| hsa-miR-181c | 1 | | | | | | |
| hsa-miR-181d | 1 | | | | | | |
| hsa-miR-185 | 1 | | | | | | |
| hsa-miR-186 | 1 | | | | | | |
| hsa-miR-18a | 1 | | | | | | |
| hsa-miR-18b | 1 | | | | | | |
| hsa-miR-190 | 1 | | | | | | |
| hsa-miR-191 | 1 | | | | | | |
| hsa-miR-192 | 1 | | | | | | |
| hsa-miR-193a-3p | 1 | | | | | | |
| hsa-miR-194 | 1 | | | | | | |
| hsa-miR-195 | 1 | | | | | | |
| hsa-miR-199a-3p/hsa-miR-199b-3p | 1 | | | Down | | | 2 |
| hsa-miR-199a-5p | 1 | | | Down | | | 2 |
| hsa-miR-199b-5p | 1 | | | | | | |
| hsa-miR-19a | 1 | | | | | Down | 2 |
| hsa-miR-19b | 1 | | | Down | Down | Down | 4 |
| hsa-miR-19b-1 | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-200a | 1 | | | | | | |
| hsa-miR-200b | 1 | | | | | | |
| hsa-miR-200c | 1 | | | | | | |
| hsa-miR-208a | 1 | | | | | | |
| hsa-miR-20a | 1 | | | | | | |
| hsa-miR-20b | 1 | | | | | | |
| hsa-miR-21 | 1 | | | | Up | | 2 |
| hsa-miR-215 | 1 | | | | | | |
| hsa-miR-219-5p | 1 | | | | | | |
| hsa-miR-22 | 1 | | | | | | |
| hsa-miR-220b | 1 | | | | | | |
| hsa-miR-221 | 1 | | | | | | |
| hsa-miR-222 | 1 | | | | | | |
| hsa-miR-223 | 1 | | | | Down | | 2 |
| hsa-miR-23a | 1 | | | | | | |
| hsa-miR-23b | 1 | | | | | | |
| hsa-miR-24 | 1 | | | | | | |
| hsa-miR-24-2 | 1 | | | | | | |
| hsa-miR-25 | 1 | | | | | | |
| hsa-miR-26a | 1 | | | Down | | | 2 |
| hsa-miR-26a-2 | 1 | | | | | | |
| hsa-miR-26b | 1 | | | | | | |
| hsa-miR-27a | 1 | | | | | | |
| hsa-miR-27b | 1 | | | | | | |
| hsa-miR-28-3p | 1 | | | | | | |
| hsa-miR-28-5p | 1 | | | Down | | | 2 |
| hsa-miR-29a | 1 | | | | Down | | 2 |
| hsa-miR-29b | 1 | | | Down | | | 2 |
| hsa-miR-29b-1 | 1 | | | | | | |
| hsa-miR-29b-2 | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-29c | 1 | | | Down | Down | | 3 |
| hsa-miR-301a | 1 | | | Down | | | 2 |
| hsa-miR-301b | 1 | | | | | | |
| hsa-miR-30a | 1 | | | | | | |
| hsa-miR-30b | 1 | | | Down | | | 2 |
| hsa-miR-30c | 1 | | | Down | | | 2 |
| hsa-miR-30d | 1 | | | | | | |
| hsa-miR-30e | 1 | | | | | | |
| hsa-miR-31 | 1 | | | | | | |
| hsa-miR-32 | 1 | | | | | | |
| hsa-miR-324-5p | 1 | | | | | | |
| hsa-miR-331-3p | 1 | | | | | | |
| hsa-miR-331-5p | 1 | | | | | | |
| hsa-miR-335 | 1 | | | Down | | | 2 |
| hsa-miR-337-5p | 1 | | | | | | |
| hsa-miR-338-3p | 1 | | | | | | |
| hsa-miR-339-3p | 1 | | | | | | |
| hsa-miR-339-5p | 1 | | | | | | |
| hsa-miR-33a | 1 | | | | | | |
| hsa-miR-340 | 1 | | | | | | |
| hsa-miR-342-3p | 1 | | | | | | |
| hsa-miR-342-5p | 1 | | | | | | |
| hsa-miR-345 | 1 | | | | | | |
| hsa-miR-34a | 1 | | | | | | |
| hsa-miR-361-3p | 1 | | | | | | |
| hsa-miR-361-5p | 1 | | | | | | |
| hsa-miR-362-3p | 1 | | | | | | |
| hsa-miR-362-5p | 1 | | | | | | |
| hsa-miR-363 | 1 | | | | | | |
| hsa-miR-369-3p | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-370 | 1 | | | | | | |
| hsa-miR-374a | 1 | | | Down | | | 2 |
| hsa-miR-374b | 1 | | | Down | | | 2 |
| hsa-miR-376a | 1 | | | | | | |
| hsa-miR-376b | 1 | | | | | | |
| hsa-miR-376c | 1 | | | | | | |
| hsa-miR-377 | 1 | | | | | | |
| hsa-miR-378 | 1 | | | | | | |
| hsa-miR-421 | 1 | | | | | | |
| hsa-miR-422a | 1 | | | | | | |
| hsa-miR-423-3p | 1 | | | | | | |
| hsa-miR-424 | 1 | | | | | | |
| hsa-miR-425 | 1 | | | | | | |
| hsa-miR-450a | 1 | | | | | | |
| hsa-miR-451 | | | | | | Down | |
| hsa-miR-454 | 1 | | | | | | |
| hsa-miR-484 | 1 | | | | Up | Up | 3 |
| hsa-miR-487b | 1 | | | | | | |
| hsa-miR-491-3p | 1 | | | | | | |
| hsa-miR-494 | 1 | | | | | | |
| hsa-miR-495 | 1 | | | | | | |
| hsa-miR-505 | 1 | | | | | | |
| hsa-miR-519d | 1 | | | | | | |
| hsa-miR-519e | 1 | | | | | | |
| hsa-miR-520d-5p | 1 | | | | | | |
| hsa-miR-524-5p | 1 | | | | | | |
| hsa-miR-532-5p | 1 | | | | Down | Down | 3 |
| hsa-miR-545 | 1 | | | | | | |
| hsa-miR-548c-5p | 1 | | | | | | |
| hsa-miR-548d-5p | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-576-5p | 1 | | | | | | |
| hsa-miR-582-5p | 1 | | | | | | |
| hsa-miR-584 | 1 | | | | | | |
| hsa-miR-589 | 1 | | | | | | |
| hsa-miR-590-5p | 1 | | | | | | |
| hsa-miR-597 | 1 | | | | | | |
| hsa-miR-598 | 1 | | | | | | |
| hsa-miR-599 | 1 | | | | | | |
| hsa-miR-600 | 1 | | | | | | |
| hsa-miR-605 | 1 | | | | | | |
| hsa-miR-606 | 1 | | | | | | |
| hsa-miR-625 | 1 | | | | | | |
| hsa-miR-627 | 1 | | | | | | |
| hsa-miR-629 | 1 | | | | | | |
| hsa-miR-654-3p | 1 | | | | | | |
| hsa-miR-656 | 1 | | | | | | |
| hsa-miR-660 | 1 | | | | | | |
| hsa-miR-7 | 1 | | | | | | |
| hsa-miR-708 | 1 | | | | | | |
| hsa-miR-7-1 | 1 | | | | | | |
| hsa-miR-720 | 1 | | | | | | |
| hsa-miR-744 | 1 | | | | | | |
| hsa-miR-769-5p | 1 | | | | | | |
| hsa-miR-770-5p | 1 | | | | | | |
| hsa-miR-876-5p | 1 | | | | | | |
| hsa-miR-887 | 1 | | | | | | |
| hsa-miR-888 | 1 | | | | | | |
| hsa-miR-9 | 1 | | | | | | |
| hsa-miR-92a | 1 | | | | Up | | 2 |
| hsa-miR-92b | 1 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| hsa-miR-93 | 1 | | | | | | |
| hsa-miR-933 | 1 | | | | | | |
| hsa-miR-95 | 1 | | | | | | |
| hsa-miR-98 | 1 | | | | | | |
| hsa-miR-99a | 1 | | | | | | |
| hsa-miR-99b | 1 | | | | | | |
| INHBA-AS1 | | | Up | | | | |
| KB-1043D8.6 | | | Up | | | | |
| KCNQ5-AS1 | | | Up | | | | |
| KCP | | | Up | | | | |
| kshv-miR-K12-5 | 1 | | | | | | |
| LARS2-AS1 | | | Up | | | | |
| LBX1-AS1 | | | Up | | | | |
| LINC00243 | | | Up | | | | |
| LINC00265 | | | Down | | | | |
| LINC00271 | | | Up | | | | |
| LINC00299 | | | Up | | | | |
| LINC00305 | | | Up | | | | |
| LINC00305 | | | Up | | | | |
| LINC00337 | | | Up | | | | |
| LINC00355 | | | Up | | | | |
| LINC00477 | | | Up | | | | |
| LINC00605 | | | Up | | | | |
| LINC00629 | | | Down | | | | |
| LINC00635 | | | Down | | | | |
| LINC00704 | | | Down | | | | |
| LINC00840 | | | Up | | | | |
| LINC00877 | | | Up | | | | |
| LINC00894 | | | Down | | | | |
| LINC00937 | | | Down | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| LINC00972 | | | Up | | | | |
| LPP-AS1 | | | Up | | | | |
| miRPlus_17891 | 1 | | | | | | |
| miRPlus_28302 | 1 | | | | | | |
| miRPlus_42720 | 1 | | | | | | |
| miRPlus_42780 | 1 | | | | | | |
| miRPlus_42859 | 1 | | | | | | |
| MSN | | | Up | | | | |
| NLGN4Y-AS1 | | | Down | | | | |
| OFD1P6Y | | | Up | | | | |
| PCA3 | | Up | | | | | |
| PCYT1B-AS1 | | | Up | | | | |
| PGM5-AS1 | | | Up | | | | |
| PTPRG-AS1 | | | Up | | | | |
| RP11-101C11.1 | | Up | | | | | |
| RP11-1023P17.2 | | | Down | | | | |
| RP11-1060J15.4 | | | Up | | | | |
| RP11-106M7.1 | | | Up | | | | |
| RP11-120C12.3 | | | Up | | | | |
| RP11-124N14.3 | | Up | | | | | |
| RP11-127L21.1 | | | Up | | | | |
| RP11-130C19.3 | | | Up | | | | |
| RP11-142G7.2 | | | Up | | | | |
| RP11-160H12.2 | | | Up | | | | |
| RP11-161I10.1 | | | Up | | | | |
| RP11-162G9.1 | | | Up | | | | |
| RP11-166N17.1 | | | Down | | | | |
| RP11-168K9.1 | | | Up | | | | |
| RP11-180M15.4 | | | Up | | | | |
| RP11-199O14.1 | | | Down | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RP11-1L12.3 | | | Down | | | | |
| RP11-203B9.4 | | | Down | | | | |
| RP11-203L2.4 | | | Down | | | | |
| RP11-217L21.1 | | | Up | | | | |
| RP11-246A10.1 | | | Up | | | | |
| RP11-24P14.1 | | | Up | | | | |
| RP11-272P10.2 | | | Up | | | | |
| RP11-274B18.4 | | | Up | | | | |
| RP11-27P7.1 | | | Up | | | | |
| RP11-281O15.4 | | | Up | | | | |
| RP11-282K24.3 | | | Up | | | | |
| RP11-292B1.2 | | | Up | | | | |
| RP11-307P5.1 | | | Down | | | | |
| RP11-315I20.1 | | | Up | | | | |
| RP11-31K23.2 | | | Down | | | | |
| RP11-328K4.1 | | | Down | | | | |
| RP11-32D16.1 | | | Up | | | | |
| RP11-331K21.1 | | | Up | | | | |
| RP11-337A23.6 | | | Down | | | | |
| RP11-342C24.8 | | | Up | | | | |
| RP11-342D11.3 | | | Down | | | | |
| RP11-351J23.1 | | | Down | | | | |
| RP11-357F12.1 | | | Up | | | | |
| RP11-358M14.2 | | | Up | | | | |
| RP11-360P21.2 | | | Down | | | | |
| RP11-386B13.3 | | | Up | | | | |
| RP11-390E23.3 | | | Down | | | | |
| RP11-398K22.12 | | | Down | | | | |
| RP11-399H11.2 | | | Up | | | | |
| RP11-3K16.2 | | | Up | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RP11-409K20.6 | | Down | | | | | |
| RP11-425I13.3 | | Up | | | | | |
| RP11-431K24.4 | | | Up | | | | |
| RP11-460I13.2 | | | Up | | | | |
| RP11-462G22.1 | | Up | | | | | |
| RP11-467I20.6 | | | Up | | | | |
| RP11-473O4.4 | | | Up | | | | |
| RP11-475O6.1 | | | Up | | | | |
| RP11-479J7.1 | | | Down | | | | |
| RP11-491H19.1 | | | Up | | | | |
| RP11-497K21.1 | | | Up | | | | |
| RP11-501O2.3 | | | Up | | | | |
| RP11-509J21.1 | | | Up | | | | |
| RP11-523O18.1 | | | Down | | | | |
| RP11-527D15.1 | | | Up | | | | |
| RP11-533O20.2 | | Down | | | | | |
| RP11-542K23.9 | | Down | | | | | |
| RP11-544M22.1 | | | Down | | | | |
| RP11-560I19.1 | | | Up | | | | |
| RP11-567N4.3 | | | Up | | | | |
| RP11-574K11.5 | | | Up | | | | |
| RP11-574K11.8 | | | Down | | | | |
| RP11-57H14.3 | | | Up | | | | |
| RP11-5N11.2 | | | Up | | | | |
| RP11-624D11.2 | | | Up | | | | |
| RP11-627G23.1 | | | Up | | | | |
| RP11-646J21.6 | | | Up | | | | |
| RP11-664D1.1 | | | Up | | | | |
| RP11-673E1.1 | | | Down | | | | |
| RP11-686G23.2 | | | Up | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RP11-702F3.3 | | | Up | | | | |
| RP11-716D16.1 | | | Up | | | | |
| RP11-729I10.2 | | | Up | | | | |
| RP11-779P15.2 | | | Up | | | | |
| RP11-782C8.1 | | | Down | | | | |
| RP11-79P5.3 | | Up | Up | | | | 2 |
| RP11-83J21.3 | | | Up | | | | |
| RP11-844P9.1 | | | Up | | | | |
| RP11-909N17.2 | | | Up | | | | |
| RP11-91K9.1 | | | Up | | | | |
| RP11-978I15.10 | | | Up | | | | |
| RP11-98J9.1 | | | Up | | | | |
| RP1-207H1.3 | | | Up | | | | |
| RP1-23E21.2 | | | Up | | | | |
| RP1-23K20.2 | | | Up | | | | |
| RP1-249H1.2 | | | Up | | | | |
| RP1-290I10.7 | | | Up | | | | |
| RP13-297E16.4 | | | Up | | | | |
| RP13-436F16.1 | | | Up | | | | |
| RP13-507P19.2 | | Up | Down | | | | |
| RP3-332B22.1 | | | Up | | | | |
| RP4-598G3.1 | | | Up | | | | |
| RP4-633O19__A.1 | | | Up | | | | |
| RP4-705O1.1 | | Down | | | | | |
| RP4-735C1.4 | | | Up | | | | |
| RP4-781K5.6 | | | Up | | | | |
| RP4-794H19.1 | | | Down | | | | |
| RP5-1091N2.9 | | | Up | | | | |
| RP5-1160K1.3 | | | Up | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RP5-1185I7.1 | | | Up | | | | |
| RP5-850O15.3 | | | Up | | | | |
| RP5-857K21.4 | | | Up | | | | |
| RP5-991C6.4 | | | Up | | | | |
| RP6-1O2.1 | | | Down | | | | |
| RP6-24A23.3 | | | Up | | | | |
| RRP7B | | | Up | | | | |
| SCAMP1 | | | Down | | | | |
| SEC62-AS1 | | | Up | | | | |
| SOS1-IT1 | | | Up | | | | |
| STX18-AS1 | | | Up | | | | |
| STX18-IT1 | | | Up | | | | |
| TEX38 | | | Up | | | | |
| TFAP2A-AS1 | | | Up | | | | |
| THRB-AS1 | | | Up | | | | |
| TTN-AS1 | | | Up | | | | |
| TTTY10 | | | Up | | | | |
| U1 | | Up | | | | | |
| UBA6-AS1 | | | Up | | | | |
| UBOX5-AS1 | | | Up | | | | |
| WASIR2 | | | Down | | | | |
| WWTR1-AS1 | | | Down | | | | |
| XXbac-BPG13B8.10 | | | Up | | | | |
| ZNRD1-AS1 | | | Up | | | | |

# Appendix V

## Genome wide mRNA and miRNA gene expression studies

*Avaliable from author

# Appendix VI

## Consent Form

Version 5 Date: 26/10/2015, Shared roots Informed consent document

# PARTICIPANT INFORMATION AND INFORMED CONSENT FORM FOR RESEARCH INVOLVING GENETIC STUDIES

**RESEARCH PROJECT:**
**Understanding the SHARED ROOTS of Neuropsychiatric Disorders and Modifiable Risk Factors for Cardiovascular Disease**

**STUDY REFERENCE NUMBER: N13/08/115**

**PRINCIPAL INVESTIGATOR:** Prof Soraya Seedat

**ADDRESS:** Department of Psychiatry,
Faculty of Medicine and Health Sciences
Stellenbosch University
PO Box 19063
Tygerberg
Cape Town

**CONTACT NUMBERS:**
Tel: 021 938 9374 (Prof Soraya Seedat)
Tel: 021 938 9207 (Health Research Ethics Committee at Stellenbosch University)
Tel: 021 938 9228 / 9768 (Tygerberg study offices)
Tel: 021 910 3605 (Stikland study unit)

We would like to invite you to participate in a research study that involves genetic analysis and possible long-term storage of blood and tissue specimens. Please take some time to read the information presented here which will explain the details of this project. Please ask the study staff or doctor any questions about any part of this project that you do not fully understand. It is very important that you are fully satisfied that you clearly understand what this research entails and how you could be involved. Also, your participation is **entirely voluntary** and you are **free to decline to participate**. If you say no, this will not affect you negatively in any way whatsoever. You are also **free to withdraw from the study at any point**, even if you do agree to take part initially.

This research study has been approved by the ethics **Health Research Ethics Committee at Stellenbosch University** and it will be conducted according to international and locally accepted ethical guidelines for research, namely the Declaration of Helsinki, and the SA Department of Health's 2004 Guidelines: *Ethics in Health Research: Principles, Structures and Processes.*

**What is genetic research?**
Genetic material, also called DNA or RNA, is usually obtained from a small blood sample. Occasionally genetic material is obtained from other sources such as saliva or biopsy specimens. A biopsy is a tiny piece of tissue that is cut out e.g. from the skin or from a lump, to help your doctor make a diagnosis. Genes are found in every cell in the human body. Our genes determine what we look like and sometimes what kind of diseases we may be susceptible to.

Worldwide, researchers in the field of genetics are continuously discovering new information that may be of great benefit to future generations and also that may benefit people today, who suffer from particular diseases or conditions.

**What does this particular research study involve?**
In a nutshell, this research project will try to identify the genes and disease pathways that cause Parkinson's disease, posttraumatic stress disorder, schizophrenia and metabolic syndrome. Metabolic syndrome is a cluster of features such as increased blood pressure, a high blood sugar level, excess body fat around the waist and abnormal cholesterol levels that occur together and increase your risk of heart disease, stroke and diabetes.

Certain blood tests can help us to recognize individuals who have a disease versus those who don't have that particular disease. These tests are often called biomarkers as they are biological markers of a particular disease. Some of these biomarkers are well known, like doing a fasting blood glucose test to see if someone has diabetes. Scientists are constantly searching for new biomarkers that can help us to better understand diseases. These biomarkers are mostly used in experiments and can't yet be used to tell if someone has a disease or not. They however help us to better understand the diseases we study and to be able to improve healthcare in the future. We want to obtain blood samples from all participants to perform biomarker testing. Examples of biomarkers that can be measured in blood include hormones, proteins and enzymes.

The causes of these brain disorders and metabolic syndrome are currently not well-understood. Some scientists think that there is a link between these disorders. We want to investigate this and will also try to find out whether there are specific genes and biological pathways that cause these disorders to develop. By studying these biological pathways we may understand what goes wrong in the affected tissue and this may eventually lead to better and more appropriate treatments for these conditions.

We will use a number of specialised genetic techniques to identify these genes and pathways, such as 'whole exome sequencing' and 'transcriptomics'. Whole exome sequencing will allow us to examine thousands of different genes simultaneously to look for gene defects. Transcriptomics will allow us to determine when and how these genes are switched on and off. We will also take photographs of your brain to see how it functions under different conditions.

Lastly, we will need to obtain a small piece of skin tissue from you to grow skin cells called fibroblasts. Later these fibroblasts will be used to create special cells called stem cells which are unique in that they can, under the right growth conditions, be used to grow any type of cell in the body. We will use them to grow brain cell lines as these are the cells affected by your disorder. These cell lines are like copies of your cells but can only exist in the laboratory under special growth conditions. We will use these cell lines only to investigate possible disease pathways and processes involved in Parkinson's disease, posttraumatic stress disorder, schizophrenia and metabolic syndrome. Also, we will only investigate the genes involved in these disorders.

An additional test that we would like to do is hair sampling. Cortisol is one of the hormones our bodies produce in response to stress. Sometimes our bodies can secrete too much or too little cortisol and this can have negative effects on our health. We can now measure cortisol and other hormones in hair samples. This is very useful as we can get an idea of the body's long term release of these hormones and collecting hair samples is painless and easy to do.

**Why have you been invited to participate?**
You have been invited to participate as you have been diagnosed with one of the following conditions, Parkinson's disease, post-traumatic stress disorder or schizophrenia. In addition to having one of these brain disorders you may also have metabolic syndrome. Alternatively you have been invited to participate to act as a control in the study – in other words you have none of the above diseases and this will allow us to look for differences between those who have these diseases and those who do not.

In order to determine whether there is a link between these brain disorders and metabolic syndrome we will recruit a total of 600 patients. Of these, 300 will have one of the three brain disorders *with* metabolic syndrome and the other 300 will have one of the three brain disorders and *will not have* metabolic syndrome.

**What procedures will be involved in this research?**
We will need to perform the following procedures on you:

Visit 1: After signing the consent form you will be required to complete a few questionnaires about your specific brain disorder. In females we will do a urine pregnancy test as pregnant individuals will be excluded from the study due to safety concerns. You will also be asked about your medical history and a doctor will perform a brief physical examination. This will take approximately 2-3 hours.

Visit 2: You will need to **fast the evening and morning prior** to this visit. A doctor or nurse will take 40-50ml (8-10 teaspoons) of blood from you for the genetic studies, biomarker testing and also for various laboratory tests to see if you have metabolic syndrome. These tests include measuring your blood sugar (glucose) and fats in your blood (triglycerides and cholesterol) as well as other tests related to the metabolic syndrome: C-reactive protein (CRP) and glycosylated haemoglobin (HbA1c). You will be asked to complete some additional questionnaires related to your health and usual practices, such as diet, exercise and smoking.

We will also take your blood pressure and pulse and do some body measurements (height, weight, waist, hip, upper body and neck circumference). In addition, if you agree, we will take one tiny piece of skin from you, measuring 3mm X 3mm. This will be taken from the forearm. You will not require any stitches. This procedure should take about 30 minutes.
At your second visit we also want to collect a very small sample of your hair, about half the thickness of a pencil (3mm). The hair is taken from the back of your head. We carefully collect together a small group of hairs and tie them together. The hair sample is then cut with scissors as close to the skin as possible. The hair will then be stored and later analysed.

At one of your visits you will undergo neuropsychological testing; this will involve doing tests to assess your performance in different skills, such as memory and solving problems.

Visit 3: In a subset of patients, we will require a third visit. At this visit we will take some specialised photos of your brain using a technique known as magnetic resonance imaging (MRI). You will be performing certain tasks in the scanner as well as just relaxing. Before the scanning starts you will be allowed to get used to the environment and the tasks will be explained to you. One of the tasks involves possibly being rewarded with money. This money will be given to you right after the scan is finished and is over and above the travel money you have already received. For a second task, some participants will also be asked to look at a number of pictures. We will be asking you how each picture makes you feel soon afterward. We will also collect urine samples to test for the presence of certain medications and drugs on the day of the scans. These results will only be used to assist us in seeing if these substances have any effects on the photos we take during the tasks and won't be used for any other reason. The whole scan session will take approximately 1 to 1 ½ hours. Although we do not give the scan

results out to all participants, as they are experimental scans not meant for clinical purposes, we will inform you should the scan show something that needs treatment.

Follow-up visit: In a subset of patients, 12 months after the first visit, we need to repeat some of the measures such as taking your blood pressure and measurements and completing some questionnaires again and will require you to donate 25ml (5 teaspoons) of blood to test for metabolic syndrome. We will also need to take specialised photos of your brain again. This should take 1 to 1 ½ hours.

Please note that during any of the above mentioned visits, patients will also have the opportunity, should they provide informed consent, to provide researchers with about 1 teaspoon of stool sample (in a special container). This sample will be used to analyse the microorganisms that are present in your gut. Previous studies have shown that there is an important link between the brain and the gut.

**Follow-up assessments**
We want to repeat all the assessments done initially around 12 and 24 months after your first visit. Repeating the assessments will allow us to get a better indication about how disease risk factors change over time. We will contact you to ask if you will be willing to return and to set up follow-up appointments at about 11 months and 23 months after your initial assessments.

**Are there any risks involved in this research?**
There are risks associated with all medical procedures. In this research project you may experience the following:

Visit length: Your visits will all generally be around 3 hours long, so you may get tired at times; we will thus ensure you have as many breaks during your visit as needed.

Blood taking: You may experience some pain or bruising at the site where blood is taken. We will attempt to minimise this using experienced medical practitioners.

Skin biopsy: If you agree to give us a piece of skin, there may be pain at the site where the skin is taken. The sample will be obtained by a process known as a ‘punch biopsy’. In this procedure, the skin on your arm will be cleaned and injected with a local anesthetic in order to minimize the pain. A small round blade will be used, and one small round piece of skin, about 3mm, will be removed. The wound will be covered with ointment and plaster and you will receive instructions on how to care for the wound before you leave. You will be given cream and plaster to cover the site before you leave. The wound usually heals after about 3 days. There will be a small scar at the site where the tissue was removed but this is likely to fade over time. If you experience any redness, intense irritation or yellowish discharge at the site of the biopsy please phone the study office immediately.

Hair sampling: The main risk for this test is that there may be a small area on your head where the hair was cut that may be visible until it has started growing out again. To reduce this risk we take the smallest possible sample, or if needed rather take 2-3 smaller samples to get enough hair. We also take from the back of the head where the overlying hair will tend to cover the area from which the sample was taken.

Brain imaging: You many experience claustrophobic feelings in the brain scanner, but if that occurs the scanning will be discontinued immediately at your request. When the machine is switched on, it will make some loud banging noises, but you will be clearly warned when this will take place. To minimize the possible discomfort associated with this, we will give you some soft earplugs to put in and will also put earphones on so that you can listen to music if you so

choose. Should you be performing the picture task, you might find some of the pictures disturbing. You are free to end the task at any time however, should you find the pictures too upsetting.

Stool sample: There is no risk involved in providing a stool sample.

**Are there any benefits to your taking part in this study and will you get told your results?**
If we determine that you have metabolic syndrome you and your doctor or GP ……. (Dr's name) will be informed of this finding. This is important as by making changes to your diet and lifestyle and by being more closely monitored by your doctor you could be prevented from having heart disease, a stroke or diabetes in the future. We will also provide you with some lifestyle advice that can assist you in making better choices regarding your health.

Should it become clear that certain abnormalities relating to the microorganisms in your gut can possibly be treated through changes in your diet or with probiotic treatment, those options will be discussed with you.

Findings from this study may benefit people with Parkinson's disease, post-traumatic disorder or schizophrenia in the future.

**How long will your bloods and tissue be stored and where will it be stored?**
The DNA and RNA extracted from your blood as well the cells grown from the skin tissue and the blood obtained for biomarker testing will be stored at the Stellenbosch University for 15 years.

We work with other researchers in South Africa and overseas and there is a chance that we may send the DNA, RNA and cells to their laboratories. In that event we will first get permission from Health Research Ethics Committee at Stellenbosch University.

After analysis we will store the remaining hair samples for a maximum period of 15 years to allow us to repeat analysis if needed.

**If your blood is to be stored is there a chance that it will be used for other research?**
Your DNA, RNA and cells will only be used for genetic research that is directly related to Parkinson's disease, post-traumatic disorder, schizophrenia and metabolic syndrome.

Also if we wish to use your samples for additional research in these fields we will apply for permission to do so from the Health Research Ethics Committee at Stellenbosch University.

The stem cells produced from your skin cells will not be used for any reproductive cloning purposes. This type of work is prohibited by the South African National Health Act 61/2003: 57(1).

If you do not wish your samples to be stored after this research study is completed you will have an opportunity to request that it be discarded when you sign the consent form.

**How will your confidentiality be protected?**
We will keep your personal details private and will only identify your DNA, RNA and tissue samples by a unique study number. This will be linked to your personal information on a database that will be protected by a password and will only be available to a few selected researchers. If we share your samples with another laboratory we will only refer to you by your study number.

**Will you or the researchers benefit financially from this research?**
You will not be paid to take part in this study although your travel expenses may be reimbursed for the visits and for follow-up visits at approximately R100 per visit. One of the fMRI tasks that you may be asked to perform involves being rewarded with money depending on how well you perform the task.

**Additional procedures to be followed only in participants diagnosed with schizophrenia (please note that if you participated in the EONKCS study this section does NOT apply to you and you will not receive treatment from us):**
At the end of your first study visit, you will receive Flupenthixol tablets that you will have to take orally, on a daily basis for 7 days. Thereafter you will receive your first Flupenthixol injection, followed by a Flupenthixol injection every 2 weeks for the duration of 3 months. At each visit you will be assessed by a psychiatrist to monitor your progress on the Flupenthixol injection. After 3 months the psychiatrist will refer you to your local community clinic, with a letter requesting the community health workers to proceed with your Flupenthixol injectable treatment. We will follow up on you on a regular basis by telephone contact. If you experience any problems at any time, you can contact Sr. Retha Smit on 021 940 4471 (in hours) or 082 805 8225 (after hours).
All medications may have unwanted side-effects, and your treatment has been known to cause stiffness, unusual movements, weight gain, high blood sugar, high blood fats, breast swelling and lactation (milk production) and sexual dysfunction. The possible benefits of participating are that you will receive expert care, you will receive medication free-of charge, and importantly, you will be assisting us to learn more about your illness so that we can provide improved care to you and others in future.

**Important information: In the unlikely event that this research leads to the development of a commercial application or patent, you or your family will not receive any profits or royalties**

**Declaration by participant:**
By signing below, I ……..…….. agree to take part in a genetic research study entitled 'Understanding the SHARED ROOTS of Neuropsychiatric Disorders and Modifiable Risk Factors for Cardiovascular Disease '.

I declare that:

 I have read or had read to me this information and consent form and it is written in a language with which I am fluent and comfortable.

 I have had a chance to ask questions and all my questions have been adequately answered.

 I understand that taking part in this study is voluntary and I have not been pressurised to take part.

 I have received a signed duplicate copy of this consent form for my records.

**Please initial the options you choose below:**

**Long term storage of samples**
(I have the right to receive confirmation that my request has been carried out)

- I agree that my blood, tissue, hair, DNA and RNA samples can be **stored for 15 years,** but I can choose to request at any time that my stored sample be destroyed. My sample will be identified with a special study code that **will remain linked to my name and contact details**.

**OR**

- I agree that my blood, tissue, hair, DNA and RNA samples can be **stored for 15 years** after the project is completed but that it is anonymised with all possible links to my identity removed, and that the researchers may then use it for additional research in this or a related field. Once my sample is anonymised, my rights to that sample are waivered. My sample may be shipped to another laboratory in SA or abroad to be used in other research projects in this or a related field

**OR**

- Please destroy my blood, tissue, hair, DNA and RNA samples as soon as the current research project has been completed.

**Permission for skin biopsy:**

- I **give permission** to donate two small pieces of skin for research purposes.

- I **do not give permission** to donate skin for research purposes.

**Permission for hair sample:**

- I **give permission** to donate a small sample of hair for research purposes.

- I **do not give permission** to donate hair for research purposes.

**Permission for stool sample:**

- I **give permission** to donate about 1 teaspoon of stool for research purposes.

- I **do not give permission** to donate about 1 teaspoon of stool for research purposes.

**CONSENT**

..... .....
**Signature of participant Signature of witness**
Signed at (*place*) ..... DATE: .....

**Declaration by investigator:**
I *(name)* …….. declare that:
 I explained the information in this document to ……..……..
 I encouraged him/her to ask questions and took adequate time to answer them.
 I am satisfied that he/she adequately understands all aspects of the research as discussed above.
 I did /did not use an interpreter. (*If an interpreter is used then the interpreter must sign the declaration below.*

..... .....
**Signature of investigator Signature of witness**
Signed at (*place*) .....…….. DATE: ……...……...

**Declaration by interpreter:**
I *(name)* …….. declare that:
 I assisted the investigator (*name*) …….. to explain the information in this document to (*name of participant*) ……. ……. …….. using the language medium of Afrikaans/Xhosa.
 We encouraged him/her to ask questions and took adequate time to answer them.
 I conveyed a factually correct version of what was related to me.
 I am satisfied that the participant fully understands the content of this informed consent document and has had all his/her question satisfactorily answered.

..... .....
**Signature of interpreter Signature of witness**
Signed at (*place*) ..... DATE:.....

# Appendix VII

## 1. HBB primer sequence and location of forward and reverse primers:

**HBB-Exon 1(1-142 -> 142bp)**

```
                          HBB-for                                start
     1 ACATTTGCTT CTGACACAAC TGTGTTCACT AGCAACCTCA AACAGACACC ATGGTGCATC

    61 TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG
                                                          HBB-rev
   121 TTGGTGGTGA GGCCCTGGGC AGgttggtat caaggttaca agacaggttt aaggagacca
   181 atagaaactg ggcatgtgga gacagagaag actcttgggt ttctgatagg cactgactct
```

**>HBB-Exon 2(273-495 -> 223bp)**

```
   241 ctctgcctat tggtctattt tcccaccctt agGCTGCTGG TGGTCTACCC TTGGACCCAG
   301 AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG
   361 GTGAAGGCTC ATGGCAAGAA AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC
                       HBB-mRNA-rev
   421 AACCTCAAGG GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT GCACGTGGAT
   481 CCTGAGAACT TCAGGgtgag tctatgggac gcttgatgtt ttctttcccc ttcttttcta
   541 tggttaagtt catgtcatag gaaggggata agtaacaggg tacagtttag aatgggaaac
   601 agacgaatga ttgcatcagt gtggaagtct caggatcgtt ttagtttctt ttatttgctg
   661 ttcataacaa ttgttttctt ttgtttaatt cttgctttct tttttttttct tctccgcaat
   721 ttttactatt atacttaatg ccttaacatt gtgtataaca aaaggaaata tctctgagat
   781 acattaagta acttaaaaaa aaactttaca cagtctgcct agtacattac tatttggaat
   841 atatgtgtgc ttatttgcat attcataatc tccctacttt attttctttt atttttaatt
   901 gatacataat cattatacat atttatgggt taaagtgtaa tgttttaata tgtgtacaca
   961 tattgaccaa atcagggtaa ttttgcattt gtaattttaa aaaatgcttt cttcttttaa
  1021 tatacttttt tgtttatctt atttctaata ctttccctaa tctctttctt tcagggcaat
  1081 aatgatacaa tgtatcatgc ctctttgcac cattctaaag aataacagtg ataatttctg
  1141 ggttaaggca atagcaatat ctctgcatat aaatatttct gcatataaat tgtaactgat
  1201 gtaagaggtt tcatattgct aatagcagct acaatccagc taccattctg cttttatttt
  1261 atggttggga taaggctgga ttattctgag tccaagctag gccctttttgc taatcatgtt
```

**>HBB-Exon 3(1346-1606 -> 261bp)**

```
  1321 catacctctt atcttcctcc cacagCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA
  1381 TCACTTTGGC AAAGAATTCA CCCCACCAGT GCAGGCTGCC TATCAGAAAG TGGTGGCTGG
  1441 TGTGGCTAAT GCCCTGGCCC ACAAGTATCA C^TAAGCTCGC TTTCTTGCTG TCCAATTTCT
                                        stop
  1501 ATTAAAGGTT CCTTTGTTCC CTAAGTCCAA CTACTAAACT GGGGGATATT ATGAAGGGCC
  1561 TTGAGCATCT GGATTCTGCC TAATAAAAAA CATTTATTTT CATTGC
```

# Appendix VIII

## Composition of Reagents and Chemicals Used

### Cell lysis buffer (DNA Extractions)

| | |
|---|---|
| 109.54g | Sucrose |
| 10ml | Triton X-100 |
| 476mg | MgCl2 |
| 10ml | Tris-HCl stock solution, pH 8 |

Add ddH2O to a final volume of 1 liter

### Agarose Gel (1%)

| | |
|---|---|
| 1g | SeaKem LE Agarose powder |
| 100ml | 1x SB (Sodium Boride) Buffer |
| 3µl | EtBr |

### Sodium Boride Buffer (20x)

| | |
|---|---|
| 38.137g | di-sodium tetraborate decahydrate |
| 1l | $dH_2O$ |

### Ethidium Bromide (10mg/ml)

| | |
|---|---|
| 0.2g | Ethidium Bromide |
| 20ml | $dH_2O$ |

### Cresol Red

| | |
|---|---|
| 0.1g | Cresol Red Indicator |
| 26.2ml | 0.01M NaOH |
| 223.8ml | $dH_2O$ |

# Appendix IX

## Collection of Blood in Paxgene Blood RNA tubes

Blood was collected under standard phlebotomy protocol. Prior to use the PAXgene Blood RNA Tube was at room temperature (18-25°C) and correctly labeled. Blood was drawn into a 'discard tube' if the PAXgene Blood RNA tube was the only tube to be drawn. The PAXgene Blood RNA tube was the last tube drawn. Blood was collected into the PAXgene Blood RNA tube using a blood collection set using a standard venipuncture technique. The PAXgene Blood RNA tube was held vertically below the blood donor's arm during collection (Figure 1), ten seconds was allowed to ensure complete blood draw and that the blood has stopped flowing into the tube. Once blood had been collected, the PAXgene Blood RNA tube was inverted 8 to 10 times and then stored upright at room temperature or 4°C.

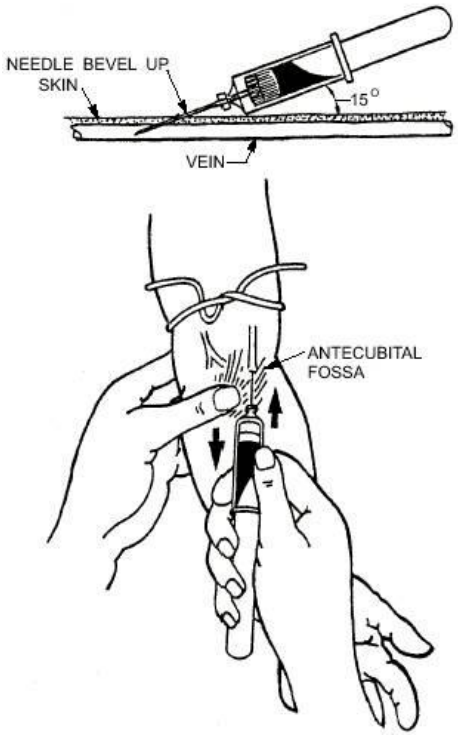



**Figure 1: Venipuncture technique for blood collection in PAXgene Blood RNA tube.**
(Figure taken from http://www.tpub.com/corpsman/227.htm)

# Appendix X

## Agilent RNA 6000 Nano Assay Protocol

Reagents and samples were equilibrated to room temperature for 30 minutes before use.

### Decontamination of the electrodes:

Electrodes were decontaminated by filling a well of an electrode cleaner with 350µl RNaseZap® and were placed into the Agilent Bioanalyser for 1 minute, following this the electrode cleaner was removed and another well of electrode cleaner was filled with 350µl RNase-free water for 10 seconds. Electrode cleaner was then removed and water on the electrodes allowed to evaporate for 10 seconds with an open lid.

### Preparation of the gel

550µl of red Agilent RNA 6000 Nano gel matrix was pipetted into the top of the spin filter and centrifuged for 10 minutes at 4000rpm. 65µl of filtered gel was aliquoted into 0.5ml RNase-free microfuge tubes (supplied with the kit) and stored at 4°C.

### Preparation of the Gel-Dye Mix

Filtered gel was equilibrated to room temperature before use. Blue RNA 6000 Nano dye concentrate was vortexed briefly and spun down; 1µl of blue RNA 6000 Nano dye was added to a 65µl aliquot of filtered gel vortexed and centrifuged for 10 minutes at room temperature at 14000rpm.

### Loading of the gel-dye mix

The RNA Nano chip was placed on the priming station and 9.0µl of gel-dye mix was added at the bottom of the well marked .The plunger was positioned at 1ml for 30 seconds and the chip priming station was closed and secured by the clip. The plunger of the syringe was pressed down until it was held by the clip and after 30 seconds the plunger was released with the clip release mechanism. The priming station was opened and 9.0µl of gel-dye mix was pipetted into the wells marked . 5µl of green RNA 6000 nano marker was pipetted into the well marked and each of the 12 sample wells.

### Loading of the ladder and samples

The ladder was thawed on ice before use to minimize secondary structure and 1µl RNA ladder was pipetted into the well labeled . 1µl sample was added into each of the 12 sample wells. The chip was placed in the IKA vortexer and vortexed for 1 minute at 2400rpm. The chip was inserted and run on the Agilent 2100 Bioanalyser.

# Appendix XI

## cDNA conversion of RNA samples

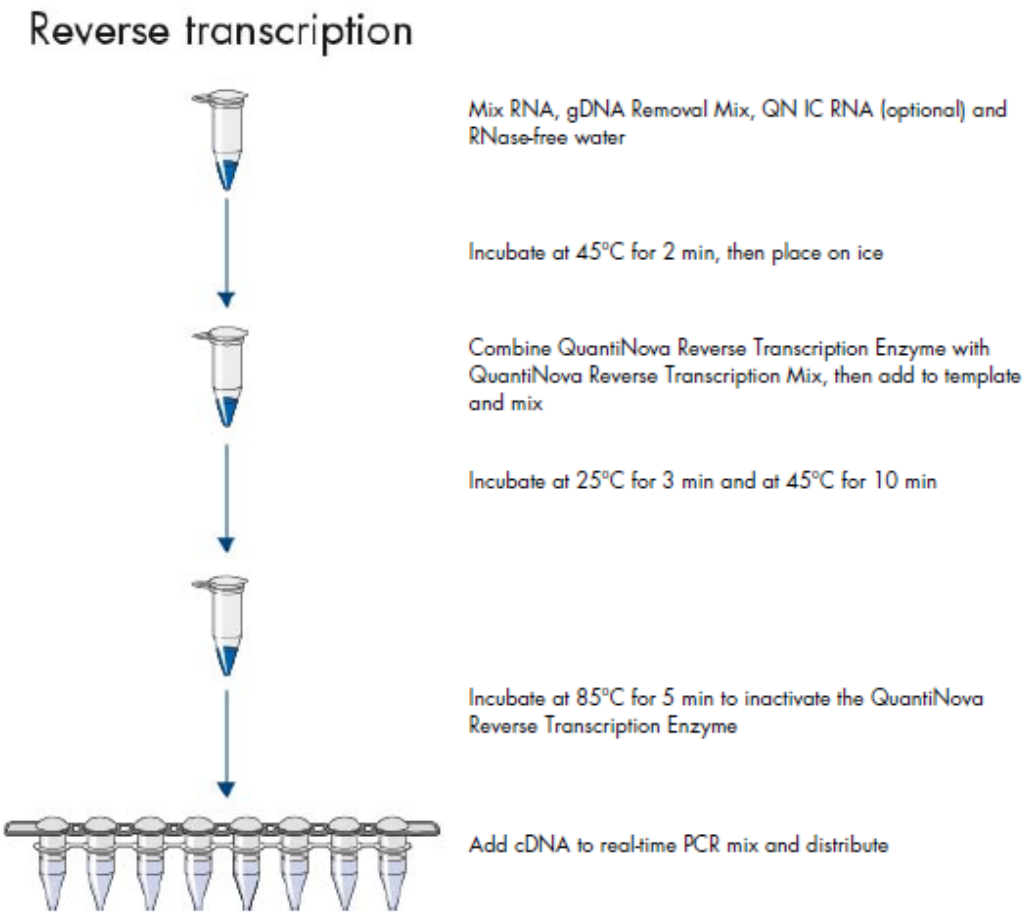



**Figure 2: Reverse transcription of RNA into cDNA using the QuantiNova™ Reverse Transcription kit (Qiagen, Hilden, Germany).**

# Appendix XII

**Table 6: RNA sequencing requirements: RIN, volumes and concentrations**

| RNA Sample | Patient or Control | RNA Concentration (ng/µl) | RIN | Volume for Sequencing (µl) | Volume Sent (RNA-Seq) (µl) | Total RNA (ng) |
|---|---|---|---|---|---|---|
| SR013 | Control | 53 | 9.6 | 14.1509434 | 15 | 795 |
| SR098 | Control | 30 | 7.8 | 25 | 25 | 750 |
| SR050 | Control | 73 | 9.7 | 10.2739726 | 15 | 1095 |
| SR079 | Control | 73 | 9.5 | 10.2739726 | 15 | 1095 |
| SR055 | Control | 119 | 9.6 | 6.302521008 | 15 | 1785 |
| SR103 | Control | 91 | 9.6 | 8.241758242 | 15 | 1365 |
| SR006 | Control | 37 | 9.3 | 20.27027027 | 20.3 | 751.1 |
| SR089 | Control | 65 | 9.3 | 11.53846154 | 15 | 975 |
| SR082 | Control | 47 | 9.6 | 15.95744681 | 16 | 752 |
| SR109 | Control | 96 | 9.6 | 7.8125 | 15 | 1440 |
| SR119 | Control | 115 | 9.9 | 6.52173913 | 15 | 1725 |
| SR099 | Control | 248 | 9.4 | 3.024193548 | 15 | 3720 |
| SR111 | Control | 184 | 9.3 | 4.076086957 | 15 | 2760 |
| SR080 | Control | 73 | 9.3 | 10.2739726 | 15 | 1095 |
| SR093 | Control | 179 | 9.3 | 4.189944134 | 15 | 2685 |
| SR088 | Control | 222 | 9.4 | 3.378378378 | 15 | 3330 |
| SR092 | Control | 243 | 8 | 3.086419753 | 15 | 3645 |
| SR054 | Control | 200 | 9.4 | 3.75 | 15 | 3000 |
| SR086 | Control | 328 | 9 | 2.286585366 | 15 | 4920 |
| SR027 | Control | 299 | 9.4 | 2.508361204 | 15 | 4485 |
| SR053 | Patient | 132 | 8.9 | 5.681818182 | 15 | 1980 |
| SR033 | Patient | 187 | 9.5 | 4.010695187 | 15 | 2805 |
| SR002 | Patient | 137 | 9.5 | 5.474452555 | 15 | 2055 |
| SR005 | Patient | 87 | 9.8 | 8.620689655 | 15 | 1305 |
| SR121 | Patient | 143 | 9.4 | 5.244755245 | 15 | 2145 |
| SR011 | Patient | 232 | 9.2 | 3.232758621 | 15 | 3480 |
| SR046 | Patient | 88 | 7.3 | 8.522727273 | 15 | 1320 |
| SR023 | Patient | 65 | 8.4 | 11.53846154 | 15 | 975 |
| SR009 | Patient | 96 | 9.1 | 7.8125 | 15 | 1440 |
| SR056 | Patient | 113 | 8.8 | 6.637168142 | 15 | 1695 |
| SR070 | Patient | 132 | 8.8 | 5.681818182 | 15 | 1980 |
| SR041 | Patient | 92 | 9.4 | 8.152173913 | 15 | 1380 |
| SR039 | Patient | 37 | 9.6 | 20.27027027 | 20.3 | 751.1 |
| SR014 | Patient | 93 | 9.4 | 8.064516129 | 15 | 1395 |
| SR042 | Patient | 220 | 8.7 | 3.409090909 | 15 | 3300 |
| SR067 | Patient | 51 | 7.9 | 14.70588235 | 15 | 765 |
| SR012 | Patient | 102 | 9.4 | 7.352941176 | 15 | 1530 |
| SR007 | Patient | 74 | 9.5 | 10.13513514 | 15 | 1110 |
| SR010 | Patient | 24 | 10 | 31.25 | 31.3 | 751.2 |
| SR003 | Patient | 114 | 9.3 | 6.578947368 | 15 | 1710 |

# Appendix XIII

## STAR RNA aligner parameters

**Table 7: STAR RNA aligner v2.3.1j default parameters**

| Parameter | Value |
|---|---|
| Generate unaligned reads | True |
| Name, sequence and quality lengths | NotEqual |
| Max buffers size | 150000000 |
| Type of flitering | Normal |
| Multimap score range | 1 |
| Max read mapping | 10 |
| Max mismatches | 10 |
| Mismatch ratio | 0.3 |
| Min Score | 0 |
| Normalized min score | 0.66 |
| Min matched bases | 0 |
| Normalized min matched bases | 0.66 |
| Filter alignment using their motifs | None |
| Collapsed splice junction reads | All |
| Max junction gap | 50000 100000 200000 |
| Non-canoncial motifs | True |
| Min overhang length for splice junctions | 30 |
| Min unique map read count per junction | 3 |
| Min total read count per junction | 3 |
| Min distance to other junctions' donor/acceptor | 10 |
| GT/AG motif | True |
| Min overhang length for splice junctions | 12 |
| Min unique map read count per junction | 1 |
| Min total read count per junction | 1 |

| | |
|---|---|
| Min distance to other junctions' donor/acceptor | 0 |
| GC/AG motif | true |
| Min overhang length for splice junctions | 12 |
| Min unique map read count per junction | 1 |
| Min total read count per junction | 1 |
| Min distnace to other junctions' donor/acceptor | 5 |
| AT/AC motif | True |
| Min overhang length for splice junctions | 12 |
| Min unique map read count per junction | 1 |
| Min total read count per junction | 1 |
| Min distance to other junctions' donor/acceptor | 10 |
| Gap open penalty | 0 |
| Non-canonical gap open penalty | -8 |
| GC/AG gap open penalty | -4 |
| AT/AC gap open penalty | -8 |
| Extra score | -0.25 |
| Deletion open penalty | -2 |
| Deletion extension penalty per base | -2 |
| Insertion open penalty | -2 |
| Insertion extension penalty per base | -2 |
| Max score reduction | 1 |
| Search start point | 50 |
| Normalized search start point | 1.0 |
| Max mapping for stitching | 10000 |
| Max seeds per read | 1000 |
| Max seeds per window | 50 |
| Max one seed loci per window | 10 |
| Max intron size | 21 |
| Max spliced alignment overhang | 5 |

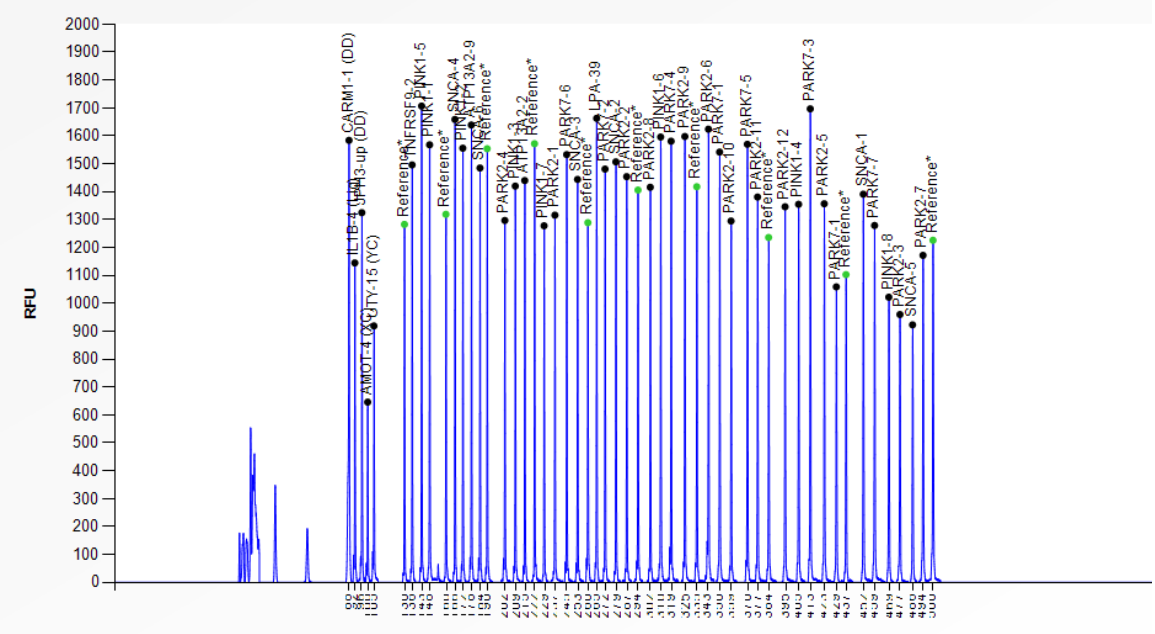
| | |
|---|---|
| Min annotated spliced alignment overhang | 3 |
| Spliced mate min read length | 0 |
| GC/AG gap open penalty | -4 |
| AT/AC gap open penalty | -8 |
| Extra score | -0.25 |
| Deletion open penalty | -2 |
| Deletion extension penalty per base | -2 |
| Insertion open penalty | -2 |
| Insertion extension penalty per base | -2 |
| Max score reduction | 1 |
| Search start point | 50 |
| Normalized search start point | 1.0 |
| Max mapping for stitching | 10000 |
| Max seeds per read | 1000 |
| Max reads per window | 50 |
| Max one seed loci per window | 10 |
| Min intron size | 21 |
| Min spliced alignment overhang | 5 |
| Min annotated spliced alignement overhang | 3 |
| Spliced mate min read length | 0 |
| Normalized spliced mate min read length | 0.66 |
| Max windows per read | 10000 |
| Max transcripts per window | 100 |
| Max hits | 10000 |
| Extra alignment score | 2 |
| Max loci anchors | 50 |
| Bin size for windows/clustering | 16 |
| Max bins between two anchors | 9 |
| Left and right flanking region size | 4 |

| | |
|---|---|
| Chimeric alignment | False |
| Cufflinks-like strand field flag | intronMotif |
| Add to quality score | 0 |
| Convert to genomic coordinates | True |
| Allow false positives | True |

# Appendix XIV

## MLPA Electropherograms

**A**




**B**

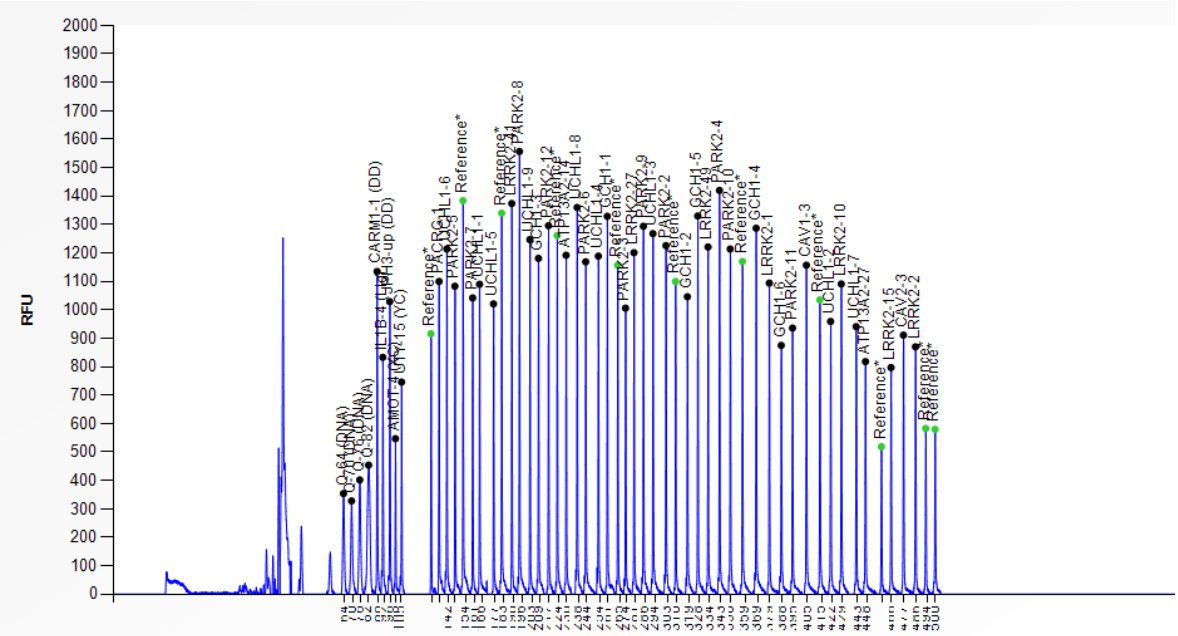
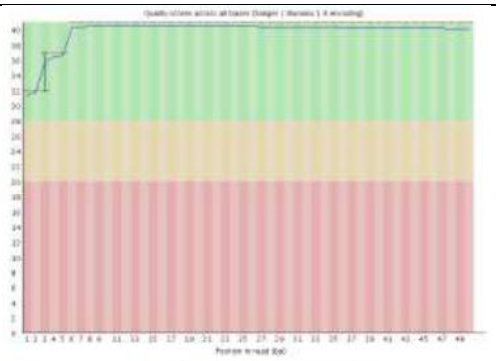
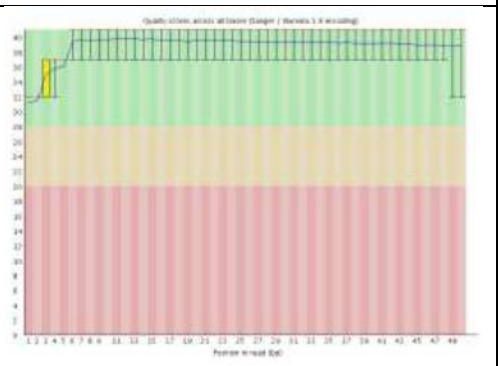
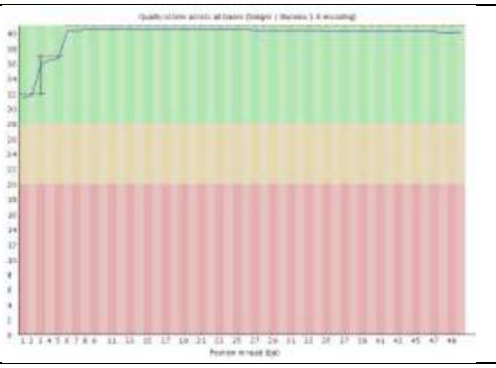
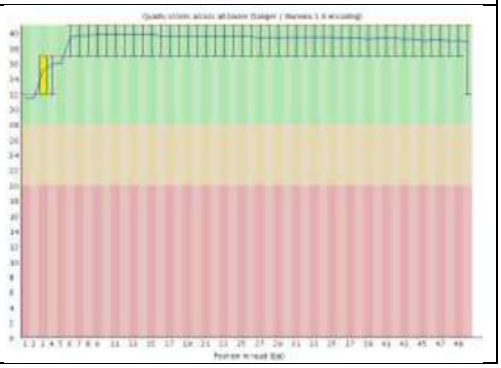



**Figure 1: A representative electropherogram using multiplex ligation-dependant probe amplification (MLPA).** Each probe for each exon is represented by a specific data point along the x-axis using the MLPA Mix 1 P051 (A) and Mix 2 P052 (B). Q-fragments are shown in B at 67, 70, 76 and 80 nt. RFU, relative fluorescence units

# Appendix XV

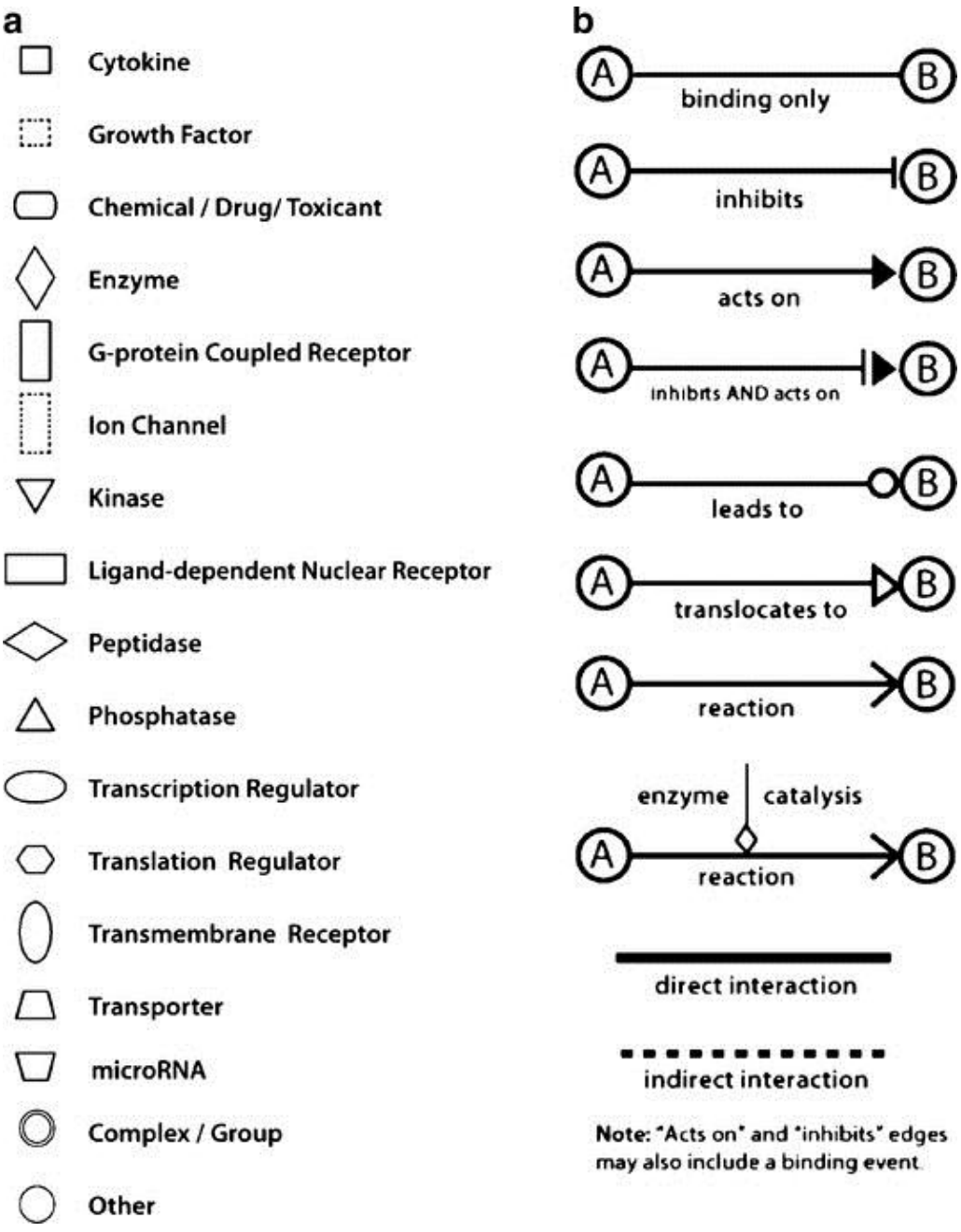
## FastQC Quality Scores

**Table 8: A representative case and control quality score across all bases using FastQC**

| Sample | Quality score across all bases Forward | Quality score across all bases Reverse |
|---|---|---|
| Control |  |  |
| Case |  |  |

# Appendix XVI

## IPA legend for canonical pathways



# Appendix XVII

## Verification using qPCR for *CEBPA*

**Table 9: Cq values for each case and control sample run in triplicate from the Lightcycler® 96 software.**

| Lab ID | *CEBPA* | *HBB* | *GAPDH* | *ACTB* |
|---|---|---|---|---|
| **Cases** | | | | |
| SR053 | 36.77 | 19.69 | 30.73 | 23.58 |
| | 38.1 | 19.69 | 30.21 | 23.33 |
| | 35.65 | 19.61 | 30.02 | 23.26 |
| SR033 | 33.71 | 17.75 | 27.76 | 21.13 |
| | 33.05 | 17.86 | 27.72 | 21.07 |
| | 33.69 | 17.95 | 27.65 | 21.18 |
| SR002 | 35.64 | 19.42 | 30.02 | 25.48 |
| | 35.75 | 19.6 | 30.21 | 25.4 |
| | 36 | 19.64 | 30.09 | 25.22 |
| SR005 | 31.73 | 17.21 | 27.07 | 23.35 |
| | 31.67 | 17.85 | 27.13 | 23.45 |
| | 31.84 | 16.92 | 26.81 | 23.4 |
| SR121 | 33.76 | 15.93 | 32.88 | 21.61 |
| | 33.54 | 15.94 | 27.76 | 21.7 |
| | 33.71 | 16.16 | 27.76 | 21.87 |
| SR011 | 34.62 | 17.34 | 28.6 | 22.54 |
| | 33.98 | 16.87 | 28.7 | 22.13 |
| | 34 | 18 | 28.97 | 22.16 |
| SR046 | 35.76 | 18.71 | 29.22 | 25.18 |
| | 34.89 | 18.58 | 29.2 | 24.98 |
| | 36.67 | 18.69 | 28.82 | 24.93 |
| SR023 | 35.63 | 20.11 | 32.99 | 26.25 |
| | 35.73 | 20.29 | 33.74 | 26.14 |
| | 35.9 | 20.25 | 32.8 | 26.2 |
| SR009 | 33.86 | 17.02 | 27.78 | 23.04 |
| | 34.13 | 17.05 | 27.71 | 23.09 |
| | 36.08 | 17.39 | 27.72 | 23.21 |
| SR056 | 35.64 | 19.42 | 30.02 | 25.48 |
| | 35.75 | 19.6 | 30.21 | 25.4 |
| | 36 | 19.64 | 30.09 | 25.22 |
| SR070 | 31.73 | 17.21 | 27.07 | 23.35 |
| | 31.67 | 17.85 | 27.13 | 23.45 |
| | 31.84 | 16.92 | 26.81 | 23.4 |
| SR041 | 35.62 | 20.49 | 32.17 | 25.82 |
| | 35.8 | 20.54 | 32.48 | 25.72 |
| | 36.41 | 20.66 | 31.92 | 25.65 |

| | | | | |
|---|---|---|---|---|
| SR039 | 35 | 23.11 | 29.63 | 27.64 |
| | 38.22 | 21.72 | 29.82 | 27.5 |
| | 33.99 | 21.76 | 31.04 | 27.47 |
| SR014 | 35.76 | 17.24 | 28.36 | 21.99 |
| | 33.65 | 17.21 | 28.41 | 21.98 |
| | 33.82 | 17.96 | 28.44 | 21.98 |
| SR042 | 35.94 | 17.5 | 29.07 | 21.55 |
| | 34.92 | 17.45 | 29.13 | 21.31 |
| | 34.69 | 17.48 | 29.03 | 21.6 |
| SR067 | 41.3 | 18.87 | 29.98 | 22.59 |
| | 36 | 18.9 | 30.11 | 23.75 |
| | 35.82 | 18.96 | 34.89 | 23.79 |
| SR012 | 34.29 | 17.95 | 28.32 | 21.95 |
| | 33.56 | 17.99 | 27.86 | 22.06 |
| | 33.89 | 18.1 | 28.63 | 22.1 |
| SR007 | 34.16 | 16.72 | 28.84 | 23.12 |
| | 34.27 | 16.68 | 28.75 | 24.28 |
| | 35.06 | 16.83 | 28.7 | 23.26 |
| SR010 | 35.63 | 18.66 | 28.81 | 22.87 |
| | 34 | 19.58 | 28.88 | 24.79 |
| | 35.01 | 18.75 | 29.69 | 23.06 |
| SR003 | 38.96 | 20.31 | 29.41 | 22.53 |
| | 37 | 19.68 | 29.39 | 22.67 |
| | 34.78 | 19.8 | 29.98 | 22.52 |
| **Controls** | | | | |
| SR013 | 30.02 | 15.87 | 25.45 | 21.52 |
| | 29.24 | 15.7 | 25.09 | 21.23 |
| | 28.91 | 15.52 | 26.05 | 21.28 |
| SR098 | 35 | 19.01 | 30.62 | 24.76 |
| | 35.03 | 19.01 | 30.35 | 24.81 |
| | 35 | 19.08 | 30.46 | 24.75 |
| SR050 | 30.02 | 15.87 | 25.45 | 21.52 |
| | 29.24 | 15.7 | 25.09 | 21.23 |
| | 28.91 | 15.52 | 26.05 | 21.28 |
| SR079 | 29 | 15.28 | 24.74 | 21.22 |
| | 28.76 | 15.68 | 24.87 | 21.11 |
| | 28.63 | 15.39 | 24.66 | 21.38 |
| SR055 | 29 | 15.28 | 24.74 | 21.22 |
| | 28.76 | 15.68 | 24.87 | 21.11 |
| | 28.63 | 15.39 | 24.66 | 21.38 |
| SR103 | 37.13 | 19.1 | 29.71 | 24.19 |
| | 36 | 18.97 | 29.96 | 24.31 |
| | 35.73 | 19.04 | 29.64 | 24.12 |

| | | | | |
|---|---|---|---|---|
| SR006 | 34.8 | 21.08 | 30.24 | 26.3 |
| | 34.6 | 21.18 | 30.41 | 25.91 |
| | 34.2 | 21.27 | 30.42 | 25.43 |
| SR089 | 36.92 | 21.1 | 32.29 | 25.52 |
| | 35.81 | 21.3 | 31.8 | 25.42 |
| | 35.9 | 21.46 | 31.8 | 25.09 |
| SR082 | 33.47 | 19.56 | 29.11 | 25.47 |
| | 33.2 | 20.57 | 28.61 | 25.73 |
| | 33.13 | 19.69 | 28.8 | 25.71 |
| SR109 | 34.18 | 20.15 | 29.63 | 24.38 |
| | 35 | 20.12 | 29.88 | 24.39 |
| | 35.71 | 20.14 | 29.56 | 24.06 |
| SR119 | 35.67 | 22.55 | 31.3 | 25.57 |
| | 40.96 | 22.58 | 31.22 | 25.53 |
| | 38 | 22.66 | 31.08 | 25.24 |
| FB | 37.05 | 29.22 | 28.67 | 23.99 |
| | 37.06 | 29.28 | 28.63 | 24.59 |
| | 37.1 | 29.02 | 28.7 | 23.75 |
| SR099 | 34.9 | 19.63 | 29.7 | 24.46 |
| | 35 | 19.58 | 29.81 | 24.23 |
| | 34.75 | 19.52 | 29.73 | 24.39 |
| SR111 | 35.75 | 18.58 | 29.99 | 23.96 |
| | 37.11 | 18.56 | 29.75 | 24.14 |
| | 35.8 | 18.64 | 29.84 | 24.06 |
| SR080 | 33.8 | 17.24 | 27.8 | 30.74 |
| | 36.18 | 17.34 | 28.05 | 20.95 |
| | 34.25 | 17.29 | 27.25 | 21.29 |
| SR093 | 34.72 | 17.7 | 28.94 | 21.46 |
| | 34.15 | 17.68 | 28.91 | 21.95 |
| | 35.68 | 17.72 | 28.83 | 21.53 |
| SR088 | 31.25 | 17.01 | 26.97 | 22.79 |
| | 32.17 | 16.34 | 26.86 | 22.57 |
| | 31.72 | 16.86 | 26.9 | 22.55 |
| SR027 | 35.76 | 19.02 | 29.12 | 22.75 |
| | 36.83 | 18.88 | 29.07 | 23.72 |
| | 35.84 | 19.05 | 28.9 | 22.9 |
| SR054 | 34.08 | 18.9 | 28.83 | 21.44 |
| | 34.26 | 17.63 | 28.84 | 22.02 |
| | 34.18 | 17.59 | 28.99 | 21.56 |
| SR086 | 34.25 | 15.96 | 28.83 | 22.11 |
| | 33.37 | 16.88 | 29.11 | 22.42 |
| | 35.76 | 16.1 | 29.31 | 22.38 |

| | | | | |
|---|---|---|---|---|
| SR092 | 33.84 | 16.19 | 29.14 | 22.56 |
| | 34.08 | 16.23 | 28.52 | 22.26 |
| | 34.26 | 17.26 | 28.58 | 22.62 |

Lightcycler® 96 software version 1.1 (Roche Diagnostics, Indianapolis, USA), FB, foetal brain; *CEBPA*, CCAAT/enhancer binding protein alpha; *HBB*, hemoglobin subunit beta; *GAPDH*, Glyceraldehyde 3-phosphate dehydrogenase; *ACTB*, β-actin